New Zealand No. 328434

International No. PCT/

TO BE ENTERED AFTER
ACCEPTANCE AND PUBLICATION

Priority dates: 25.07.1997

Complete Specification Filed: 25.07.1997

Classification:(6) C12N9/24; A01H7/00; C12N15/55,82; A61K38/47

Publication date: 27 May 1998

Journal No.: 1428

NEW ZEALAND

PATENTS ACT 1953

COMPLETE SPECIFICATION

Title of Invention:

Coniferin beta-glucosidase cdna for modifying lignin content in plants

Name, address and nationality of
applicant(s) as in international
application form:
THE UNIVERSITY OF BRITISH COLUMBIA, C/o University-Industry Liaison Office, 2194 Health Sciences Mall, Room 331 IRC Building, Vancouver BC V6T 1Z3, Canada

328434

NEW ZEALAND

PATENTS ACT, 1953

No:

Date:

COMPLETE SPECIFICATION

CONIFERIN BETA-GLUCOSIDASE cDNA FOR MODIFYING LIGNIN CONTENT IN PLANTS

We, JOHN E CARLSON, a citizen of Canada of 12064 57A Avenue, Surrey, BC V3X2S3, Canada, D PALITHA DHARMAWARDHANA, a citizen of Sri Lanka of 109 N. Cayuga Street, #F, Ithaca, New York 14850, United States of America, CARL J DOUGLAS, a citizen of Canada of 3744 Heather Street, Vancouver BC V5Z3L2, Canada, BRIAN E ELLIS, a citizen of Canada of 3771 West 38th Avenue, Vancouver, BC V6N 2Y3 and THE UNIVERSITY OF BRITISH COLUMBIA, c/o University-Industry Liaison Office, 2194 Health Sciences Mall, Room 331 IRC Building, Vancouver BC V6T 1Z3, Canada, do hereby declare the invention for which we pray that a patent may be granted to us, and the method by which it is to be performed, to be particularly described in and by the following statement:

(followed by page 1A)

CONIFERIN BETA-GLUCOSIDASE CDNA FOR MODIFYING LIGNIN CONTENT IN PLANTS

Technical Field

This invention relates to DNA molecules cloned from plants and methods of using such DNA molecules to produce transgenic plants with altered lignin content.

BACKGROUND

Lignin is the second most abundant organic material in the biosphere, and is a major component of cell walls of woody plants (such as poplar and pine species) and fodder crops (such as maize, wheat and barley). The quantity of lignin in plant material affects characteristics that are agronomically important. For example, in fodder crops the amount of lignin present determines how easily the crop may be digested by animals; relatively small increases in lignin content may produce a large decreases in the digestibility of the crop. Therefore, reducing lignin content would enhance digestibility, facilitating a more efficient use of such crops. In the timber industry, producing wood pulp for papermaking requires the removal of lignin to release the cellulosic content of the timber. The process of removing the lignin consumes large amounts of energy and produces environmentally harmful lignin waste liquors which must be treated prior to disposal. It has also been suggested that residual lignin in paper pulp may produce

toxic polychlorinated biphenols when the lignin interacts with chlorine used in the bleaching process. Thus, decreasing lignin content in wood products would be advantageous for papermaking. On the other hand, increasing the lignin content of timber offers the possibility of increased wood strength.

Accordingly, modification of quality and quantity of lignin in plants has been a long-standing interest among breeders and, more recently, among molecular biologists. Recent molecular approaches towards methods for reducing lignin content in plants are typified by: U.S. Patent No. 5,451,514, "Modification of Lignin Synthesis in Plants"; Canadian Patent No. 2,005,597, "Plants Having Reduced Lignin or Lignin of Altered Quality"; and International Patent Application Publication No. WO 94/23044.

Lignin is a complex polymer of three cinnamyl alcohols, p-coumaryl, coniferyl and sinapyl, all products of phenylpropanoid metabolism. Depending on the plant species or tissue, the relative proportion of the different monomers in lignin can vary significantly. In gymnosperms for example, lignin is predominantly composed of coniferyl alcohol monomer units, whereas angiosperms have significant proportions of sinapyl moieties. The metabolism of lignin production involves many intermediates, enzymatic pathways and, correspondingly, genes. Accordingly, there are several gene/enzyme targets that might be selected to manipulate lignin production through genetic engineering.

Alteration of lignin levels by antisense and sense suppression of gene expression has already been attempted for several enzymes in the phenylpropanoid pathway including PAL (Elkind *et al.* 1990), CAD (Schuch 1993; Canadian patent 2,005,597; U.S. Patent No. 5,541,514), 4CL (Lee and Douglas 1994) and COMT (WO 94/23044). However, all of these attempts to modify lignin synthesis are directed at early stages in the synthetic pathway and are therefore likely to interfere with other metabolic processes which share these intermediate steps. It is clear, for example, that interference with early steps in the phenylpropanoid pathway can have undesirable pleiotropic effects (Elkind *et al.*, 1990). In addition, modulating biosynthetic enzymes that act early in the pathway may not be effective because alternative synthetic routes may be available. A better approach to modulating lignin synthesis would be to regulate later stages in the lignin biosynthesis pathway: this would minimize or avoid pleiotropic effects and would likely provide a greater degree of effective control.

It is an object of the present invention to identify and provide a plant nucleic acid sequence that encodes an enzyme that functions late in the pathway of lignin biosynthesis. It is a further object of this invention to provide vectors containing forms of this nucleic acid sequence suitable for introduction into plants to modify the production of lignin.

SUMMARY OF THE INVENTION

In a first aspect, the present invention provides an isolated nucleic acid molecule comprising at least 15 consecutive nucleotides of the sequence shown in Seq. I.D. No. 6 and encoding a coniferin β-glucosidase enzyme.

In a further aspect, the present invention provides an isolated nucleic acid molecule which encodes a coniferin β-glucosidase enzyme and which hybridizes under condition of at least moderate stringency to the nucleotide sequence shown in Seq. I.D. No. 6.

In a further aspect, the present invention provides a coniferin β-glucosidase enzyme encoded by a nucleic acid molecule according to the invention.

In a still further aspect, the present invention provides an isolated oligonucleotide which comprises at least 15 consecutive nucleotides of the sequence shown in Seq. I.D. No. 6 or its complementary strand.

Also provided are recombinant vectors including a DNA sequence of the invention, and transgenic plants all as set forth in the accompanying claim set.

In a further aspect, the present invention provides a method of producing a plant with an altered lignin content relative to an untransformed plant of that species, comprising introducing into the plant a recombinant vector that comprises a promoter operably linked to a nucleic acid which hybridizes under conditions of moderate stringency to the sequence shown in Seq. I.D. No. 6 and which encodes a coniferin β-glucosidase enzyme.

(followed by page 4a)

In a further aspect, the present invention provides a method of producing a plant with an altered lignin content relative to an untransformed plant of that species, comprising introducing into the plant a recombinant vector that comprises a promoter operably linked to an antisense nucleic acid which, when expressed in cells of the plant, inhibits the expression of a native coniferin β-glucosidase gene.

In yet a further aspect, the present invention provides a method of producing a plant with an altered lignin content relative to an untransformed plant of that species, comprising introducing into the plant a nucleic acid molecule comprising a coding sequence operably linked to a promoter sequence, wherein the coding sequence encodes an untranslatable plus-sense transcript that shares at least 80% sequence similarity with a transcript of a native coniferin β-glucosidase gene.

In a still further aspect, the present invention provides an isolated nucleic acid which encodes a coniferin β-glucosidase.

In another aspect, the present invention provides a method of isolating a nucleotide sequence encoding a coniferin β-glucosidase enzyme, the method comprising hybridizing a nucleotide preparation with a DNA molecule comprising at least 15 consecutive nucleotides of the sequence set forth in Seq. I.D. No. 6.

BRIEF DESCRIPTION OF THE INVENTION

The inventors have determined that the gene encoding coniferin β-glucosidase would be an excellent target gene for modifying lignin content in plants, particularly in trees such as conifers. The coniferin β-glucosidase enzyme catalyzes the hydrolysis of the 4-O-glucoside of coniferyl alcohol, coniferin, which is one of the last steps in the biosynthesis of lignin. Thus, the level of coniferin β-glucosidase activity directly affects lignin synthesis and, therefore, the quantity of lignin in the plant tissue. Coniferin accumulates in conifer xylem during cambium reactivation, consistent with a role as the dominant lignin precursor in these species (Freudenberg and Harkin 1963, Savidge 1989). β-glucosidases capable of hydrolyzing coniferin have been detected in suspension culture systems (Hosel *et al.* 1982, Hosel and Todenhagen 1980) and seedlings (Marcinowski and Grisebach 1978), and a coniferin β-glucosidase has been purified from differentiating xylem in trees (Dharmawardhana et al., 1995). However, to date, the genetic manipulation of coniferin β-glucosidase has not been possible because the gene encoding the enzyme has not been cloned.

To that end, the inventors have cloned and sequenced a complementary DNA (cDNA) sequence from the conifer tree species *Pinus contorta*. The provision of this cDNA sequence enables, for the first time, the regulation of coniferin β-glucosidase activity in plants through genetic engineering. Specifically, the

(followed by page 5)

invention provides genetic constructs, such as plant transformation vectors, that include various forms of the coniferin β-glucosidase cDNA or sequences that are homologous to this cDNA. Depending on the specific nature of these constructs, they may be introduced into plants in order to increase or reduce the production of the coniferin β-glucosidase enzyme, and therefore to regulate lignin synthesis.

Transformation vectors according to this invention preferably include a recombinant DNA sequence that comprises all or part of the coniferin β-glucosidase cDNA. Depending on the nature of the promoter sequence selected, such constructs may be used to modify lignin content throughout the plant or in a tissue-specific manner and either constitutively or at certain stages of plant development. The availability of inducible plant promoters also offers the possibility of changing lignin biosynthesis in a plant at desired times by application of the chemical or physical agent that induces transcription from the promoter.

In one embodiment, transformation vectors may be constructed to over-express the coniferin β-glucosidase enzyme ("sense" orientation). Enhanced lignin synthesis may be achieved by introducing such vectors into plants. Examples of the application of this approach to modify plant phenotypes include U.S. Patent No. 5,268,526, "Overexpression of Phytochrome in Transgenic Plants", U.S. Patent No. 4,795,855, "Transformation and Foreign Gene Expression in Woody

Species", and U.S. Patent No. 5,443,974 (over-expression of stearoyl-ACP desaturase gene).

Alternatively, such over-expression vectors may be used to suppress coniferin β-glucosidase enzyme activity through sense-suppression, as described in U.S. Patent Nos. 5,034,323 and 5,283,184, both entitled "Genetic Engineering of Novel Plant Phenotypes".

In another embodiment, constructs may be designed to express plus-sense untranslatable coniferin β-glucosidase RNA, using methodolcgies described in U.S. Patent No. 5,583,021, "Production of Virus Resistant Plants". Constructs of this type may be used to reduce the expression of the native coniferin β-glucosidase gene, thereby reducing coniferin β-glucosidase enzyme activity and, as a result, lignin content.

In other embodiments, the present invention provides genetic constructs designed to express antisense versions of the coniferin β-glucosidase RNA. "Antisense" RNA is an RNA sequence that is the reverse complement of the mRNA encoded by a target gene. Examples of the use of antisense RNA to inhibit expression of target plant genes include U.S. Patent No. 5,451,514, "Modification of Lignin Synthesis in Plants" (use of antisense RNA to regulate CAD), U.S. Patent No. 5,356,799, "Antisense Gene Systems of Pollination Control for Hybrid Seed Production", U.S. Patent No. 5,530,192 (use of antisense RNA to alter amino acid and fatty acid composition in plants).

In conjunction with these genetic constructs,

the present invention also includes methods for altering lignin biosynthesis in plants. Generally, such methods comprise introducing into the genome of a plant a genetic construct that includes all or part of the coniferin β-glucosidase cDNA (either in sense or antisense orientation) or a sequence derived from this cDNA. Methods for introducing transformation vectors into plants are well known in the art and include electroporation of plant protoplasts, liposome-mediated transformation, polyethylene mediated transformation; transformation using viruses, micro-injection of plant cells, micro-projectile bombardment of plant cells, vacuum infiltration, and *Agrobacterium tumeficiens* (AT) mediated transformation. Methods particularly suited to the transformation of woody species are described in Ellis et al. (1993), Ellis et al. (1996), U.S. Patent No. 5,122,466, "Ballistic Transformation of Conifer" and U.S. Patent No. 4,795,855, "Transformation and Foreign Gene Expression with Woody Species".

The invention also includes transformed plants having altered lignin compositions as a result of being transformed with a genetic construct as described above. Examples of plants that may be transformed in this manner include conifers, such as plants from the genera *Picea*, *Pseudotsuga*, *Tsuga*, *Sequoia*, *Abies*, *Thuja*, *Libocedrus*, *Chamaecyparis* and *Laryx*. Pines are expected to be a particularly suitable choice for genetic modification by the methods disclosed herein, including loblolly pine (*Pinus taeda*), slash pine (*Pinus*

elliotii), longleaf pine (*Pinus palustris*), shortleaf pine (*Pinus echinata*), jack pine (*Pinus banksiana*), ponderosa pine (*Pinus ponderosa*), red pine (*Pinus resinosa*), Eastern white pine (*Pinus strobus*), Western white pine (*Pinus monticola*), sugar pine (*Pinus lambertiana*), lodgepole pine (*Pinus contorta*), Monterey pine (*Pinus radiata*), Afghan pine (*Pinus eldarica*), Scots pine (*Pinus sylvestris*) and Virginia pine (*Pinus virginiana*). Other tree species, including poplar, eucalyptus and aspen may also be transformed using the nucleotide sequences of this invention. However, the invention is not limited to trees: crop and forage plants such as maize, tobacco, alfalfa, wheat and grasses may also be transformed using the constructs provided by this invention in order to modify lignin content. In general, this invention can be applied to any plant species that can be transformed.

Throughout this specification and claims, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", is to be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 shows the nuclic acid sequence of the coniferin β-glucosidase cDNA and the amino acid sequence of the encoded protein.

Fig. 2 is a dendrogram illustrating the amino acid sequence comparisons between plant, two bacterial and one human *family 1* glycosyl hydrolases and a *family 3* glycosyl hydrolase from *Agrobacterium tumefaciens*. The dendrogram was constructed using GeneWorks CLUSTAL V program. Database accession numbers are in parentheses. 1, *A. tumefaciens* coniferin β-G (a42292); 2, *Brassica*

napus thio β-G (q00326); 3, *Sinapis alba* thio β-G (p29092); 4, *B.napus* thio β-G (s56656); 5, *B.napus* thio β-G (s39549); 6, *Arabidopsis thaliana* thio β-G (p37702); 7, *Pinus contorta* coniferin β-G; 8, *Prunus serotina* cyanogenic β-G (u50201); 9, *Prunus serotina* cyanogenic β-G (u26025); 10, *Trifolium repens* cyanogenic β-G (p26205); 11, *T.repens* β-G (26204); 12, *Costus speciosus* furostanol 26-O-β-G (d83177); 13, *Manihot esculenta* cyanogenic β-G (s23940); 14, *Oryza sativa* cyanogenic β-G (u28047); 15, *Hordeum vulgare* cyanogenic β-G (a57512); 16, *Avena sativa* β-G (s50756); 17, *Sorghum bicolor* cyanogenic β-G (u33817); 18, *Zea mays* β-G (p49235); 19, *Brassica nigra* β-G (u72154); 20, *A.thaliana* β-G (u72153); 21, *B.napus* β-G (s52771); 22, *Agrobacterium faecalis* cellobiase (g67489); 23, *Bacillus circulans* cellobiase (q03506); 24, *Homo sapiens* lactase-phlorizin hydrolase domain IV (p09848).

Fig. 3 shows the alignment of the CBG amino acid sequence (GBAA) with the following amino acid sequences: *Hordeum* β-glucosidase (L41869); *Prunus* amygladin hydrolase (U26025); *Prunus* β-glucosidase (X56733); *Trifolium* cyanogenic β-glucosidase (X56733); *Trifolium* non cyanogenic β-glucosidase (P26204); Manihot β-glucosidase (X94986); Manihot linemarase (S35175); Sorghum dhurrinase (U33817); Zea β-glucosidase (A48860); Avena β-glucosidase (X78433); Arabidopsis thioglucosidase (X89413); Brassica β-glucosidase(S52711); Brassica thioglucosidase (Q00326); Arabidopsis β-glucosidase (L11454); Human LPH subunit

III (LPH3HU); Bacillus β-glucosidase (A48969); Bacillus β-glucosidase (Q08638); Streptomyces β-glucosidase (S45675). "*" represents perfectly conserved amino acids, "." represents well conserved amino acids.

Fig. 4 shows a transformation vector suitable for introducing antisense CBG into plants.

DETAILED DESCRIPTION OF THE INVENTION

Definitions and Abbreviations

In order to facilitate review of the various embodiments of the invention, the following definitions of terms and explanations of abbreviations are provided:

4-NPG: 4-nitrophenyl β-glucoside

2-NPG: 2-nitrophenyl β-glucoside

MUG: 4-methylumbelliferyl β-glucoside

VRA-G: 5,4-(β-D-glucopyranosyloxy)-3-methoxyphenylmethylene -2-thioxothiazolidin-4-one-3-ethanoic acid. VRA-G is a substrate analog of coniferin synthesized by Biosynth International Inc., Skoke, Illinois.

EDC: 1-ethyl-3-(dimethylaminopropyl) carbodiimide

PAL: phenylalanine ammonia-lyase

CAD: Cinnamyl alcohol dehydrogenase

4CL: 4-coumarate: CoA ligase

COMT: caffeic acid 3-o-methyltransferase

PAGE: polyacrylamide gel electrophoresis

CBG: coniferin β glucosidase

Isolated: An "isolated" nucleic acid has been substantially separated or purified away from other nucleic acid sequences in the cell of the organism in which the nucleic acid naturally occurs, i.e., other chromosomal and extrachromosomal DNA and RNA. The term "isolated" thus encompasses nucleic acids purified by standard nucleic acid purification methods. The term also embraces nucleic acids prepared by recombinant expression in a host cell as well as chemically synthesized nucleic acids.

cDNA (complementary DNA): a piece of DNA lacking internal, non-coding segments (introns) and regulatory sequences which determire transcription. cDNA is synthesized in the laboratory by reverse transcription from messenger RNA extracted from cells.

ORF (open reading frame): a series of nucleotide triplets (codons) coding for amino acids without any termination codons. These sequences are usually translatable into a peptide.

Ortholog: two nucleotide sequences are orthologs of each other if they share a common ancestral sequence and diverged when a species carrying that ancestral sequence split into two species. Ortholgous sequences are also homologous sequences.

Probes and primers: Nucleic acid probes and primers may readily be prepared based on the nucleic acids provided by this invention. A probe comprises an isolated nucleic acid attached to a detectable label or reporter molecule. Typical labels include radioactive

isotopes, ligands, chemiluminescent agents, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed, e.g., in Sambrook et al. (1989) and Ausubel et al. (1987).

Primers are short nucleic acids, preferably DNA oligonucleotides 15 nucleotides or more in length. Primers may be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods known in the art.

Methods for preparing and using probes and primers are described, for example, in Sambrook et al. (1989), Ausubel et al. (1987), and Innis et al., (1990). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, MA).

Purified: the term purified does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified coniferin β-glucosidase protein preparation is one in which the coniferin β-glucosidase protein is more pure than the protein in its natural environment within a cell.

Preferably, a preparation of a coniferin β-glucosidase protein is purified such that the subject protein represents at least 50% of the total protein content of the preparation.

Operably linked: A first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein coding regions, in the same reading frame.

Recombinant: A recombinant nucleic acid is one that has a sequence that is not naturally occurring or has a sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques.

Transgenic plant: as used herein, this term refers to a plant that contains recombinant genetic material not normally found in plants of this type and which has been introduced into the plant in question (or into progenitors of the plant) by human manipulation. Thus, a plant that is grown from a plant cell into which recombinant DNA is introduced by transformation is a

transgenic plant, as are all offspring of that plant which contain the introduced DNA (whether produced sexually or asexually).

Coniferin β-glucosidase: The defining functional characteristic of the coniferin β-glucosidase enzyme is its ability to hydrolyze coniferin to release coniferyl alcohol. This activity can be measured using the glucosidase assay described herein. This invention provides a cDNA encoding the coniferin β-glucosidase enzyme from *Pinus contorta*. However the invention is not limited to this particular coniferin β-glucosidase: other nucleotide sequences which encode coniferin β-glucosidase enzymes are also part of the invention, including variants on the disclosed cDNA sequence and orthologous sequences from other plant species, the cloning of which is now enabled. Such sequences share the functional characteristic of encoding an enzyme that is capable of hydrolyzing coniferin.

Additional definitions of terms commonly used in molecular genetics can be found in Benjamin Lewin, *Genes V* published by Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew et al (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

Sequence Listing

Seq. I.D. No. 1 is primer N7A

Seq. I.D. No. 2 is primer N7B

Seq. I.D. No. 3 is primer N10

Seq. I.D. No. 4 is primer CBG172

Seq. I.D. No. 5 is primer CBG75

Seq. I.D. No. 6 is the CBG cDNA and CBG peptide

Seq. I.D. No. 7 is primer NT1

Seq. I.D. No. 8 is primer CT1

Seq. I.D. Nos. 9-12 are primers useful for amplification of the CBG cDNA sequence (see Example 4 below).

In the Sequence Listing, standard abbreviations are used for nucleotide bases, i.e., A = Adenine, G = Guanine, C = Cytosine, T = Thymine, I = Inosine, M = A or C, R = A or G, W = A or T, S = C or G, Y = C or T and K = G or T.

Detailed aspects of the invention are provided in the following examples.

EXAMPLE 1
Identification of the Coniferin β-Glucosidase cDNA

Actively differentiating *Pinus contorta* xylem was harvested as described by Dharmwardhana et al. (1995) and used to isolate total RNA as described by Lewinsohn et al. (1994). PolyA RNA isolated with an Oligotex mRNA isolation kit (Qiagen) was used to construct a cDNA library in the λZAP-XR vector,

employing Stratagene cDNA synthesis and GigapakII Gold packaging kits.

Coniferin β-glucosidase enzyme was purified from *Pinus contorta* xylem tissue as described by Dharmwardhana et al. (1995). In order to determine the N-terminal amino acid sequence of the purified enzyme, it was run on native PAGE gels, stained for activity on the synthetic coniferin substrate VRA-G and the staining band excised and subjected to SDS-PAGE. The protein was then transferred to an Immobilon membrane for N-terminal amino acid sequencing using an Applied Biosystems 470A gas phase sequencer (Edman degradation).

Gene-specific primers for PCR amplification of CBG sequence fragments were then designed based on the 15 N-terminal amino acid sequence obtained. Primers N7A and N7B were based on the first 7 N-terminal amino acid residues and were identical except at the third base from the 3' end where the degeneracy is split between the primers.

N7A: 5' GCTCTAGAGCGAC(T)A(C)GIAAC(T)AAC(T)TTTCC 3' (Seq. I.D. No. 1)

N7B: 5' GCTCTAGAGCGAC(T)A(C)GIAAC(T)AAC(T)TTCCC 3' (Seq. I.D. No. 2)

The amplification template used was the λZAP-cDNA library described above. The initial PCR reactions contained 200-300 ng λZAP-cDNA as template, 200 nM degenerate gene-specific primer N7A or N7B, 50 nM vector primer M13F or T7 (BRL), 200μM dNTP, and 1X reaction buffer (10mM TrisHCl pH 8.3, 1.5mM $MgCl_2$, 50mM KCl) in a

50μl volume. Prior to adding 3 units *Taq* polymerase (Boehringer), the reaction mixture was heated to 94°C for 2 min. The thermal cycling regime was as follows: 1-2 cycles (94°C/1min., 48°-52°C/2min., 72°/2min.); 30 cycles (94°C/45sec., 55°/1min., 72°C/2min.) ; 72°C/10min. extension.

Amplification using primer N7B yielded 3-4 major bands, whereas amplification with N7B did not yield consistent product, suggesting a mismatch at the degenerate third base. To increase specificity and identify the desired amplification product, a 20ng aliquot of reaction products from the initial PCR using N7B was reamplified using the partially nested gene-specific primer N10 [GAC(T)A(C)GIAAC(T)AAC(T) TTCCCIT(A)C(G)IGA(T)TT, Seq. I.D. No. 3] and vector primer T7 (30 cycles of 94°C/45sec., 55°/1min., 72°C/2min. followed by 72°C/10min. final extension), yielding a 1.7kb band.

Following identification of the 1.7kb band as the desired amplification product, the initial PCR reaction was repeated with less (0.9mM) $MgCl_2$ in the reaction buffer. The resulting 1.7kb band was then isolated by gel purification (Qiagen) and cloned into EcoRV-digested T-tailed Bluescript II KS vector according to the T/A cloning protocol (Holton and Graham, 1991). Plasmid minipreps from several clones were used for restriction analysis of insert and for primer-directed sequencing of both strands using ABI AmpliTaq dye termination cycle sequencing.

To amplify the 5' end of the CBG cDNA, λZAP-cDNA from the library was again used as a template, this time in conjunction with a T3 vector primer and the gene-specific primer CBG172 (CACATATCTGTGATATTGGTCG, Seq. I.D. No. 4) based on the sequence of the 3' CBG amplification product. A second nested gene-specific primer CBG75 (CCATCTTCTCGGACTGCTC, Seq. I.D. No. 5) was used to reamplify the former reaction products to confirm the authenticity of the PCR product. The cloning and sequencing of the 5' PCR product was conducted as described above. An exact sequence match in the overlapping regions of the 5' and and 3' end clones confirmed the authenticity of the 5' amplification product.

EXAMPLE 2
Analysis of the CBG cDNA Sequence

The complete CBG cDNA sequence, shown in Fig. 1 and in Seq. I.D. No. 6, is 1909 bp in length, Nucleotides 183-1721 of the 1909bp encode a 513 amino acid protein (Fig. 1). The 5' and 3'-untranslated regions of the full length sequence contain 162 and 187 nucleotides, respectively. The 3'-untranslated region does not contain the conserved eukaryotic polyadenylation signal AAUAAA, as is the case for more than 50% of reported plant mRNA sequences (Wu et al., 1995). Instead, the CBG 3'-untranslated region contains AAUAAA-like sequences like most plant mRNAs (Joshi, 1987).

The 5'-UTR of the CBG cDNA carries a 9bp AC-rich element (AACCAACAA) that is also present in *Arabidopsis* PAL1 and bean chalcone synthase (CHS15) genes, and has been proposed to be an elicitor-inducible hypersensitive site (Lawton et al., 1990; Ohl et al., 1990). This indirectly associates CBG with other phenylpropanoid metabolic genes/regulation, and is consistent with the induction of CBG activity in jackpine cell cultures by fungal elicitation (Campbell & Ellis, 1991).

The deduced 513 amino acid protein has a molecular weight of 58.3 kD and a calculated isoelectric point of pH 4.9. The N-terminal amino acid sequence determined for the purified enzyme corresponds to amino acids 24 - 40 in the deduced sequence. Met35 in the deduced sequence was identified as Thr during N-terminal amino acid sequencing. This mismatch could result from a misidentification during amino acid sequencing, or could represent a polymorphism. The nascent protein contains an N-terminal signal peptide with features characteristic of eukaryotic secretory signal sequences for ER targeting. The "weight matrix" method (von Heijne, 1986) predicts two possible cleavage sites for the signal peptide, one between residues Gly17 and Phe18, and a second between Ala23 and Arg24. Since the N-terminal amino acid sequence of the mature protein begins at Arg24, the co-translational processing of the signal peptide appears to occur at the predicted second cleavage site. The protein contains two putative N-

asparagine glycosylation sites at Asn223 and Asn447, consistent with the detection of oligosaccharide sidechains in the purified enzyme (Dharmawardhana et al., 1995).

Nucleotide and amino acid sequence homology searches and comparisons were carried out using BLAST (Altschul et al., 1990) on Genbank, EMBL, PDB, SWISS-PROT and PIR databases. Further sequence analysis was performed using PC/GENE or GeneWorks (IntelliGenetics Inc.) software. The derived amino acid sequence of CBG, when compared to other glycohydrolase sequences in the databases, showed the strongest similarity to enzymes belonging to *family 1* glycosyl hydrolases (Henrissat, 1991). The β-glucosidases showing the highest similarity (30-50% identity) to CBG were from plant species *Prunus, Hordeum, Trifolium, Manihot, Sorghum, Avena,* and *Costus*. The dendrogram in Fig 2 illustrates that among the plant β-glucosidases, pine CBG is loosely clustered with cyanogenic β-glucosidases from several species (Fig 2: sequences 7 to 13). Fig. 3 shows an alignment of the CBG amino acid sequence with other of β-glucosidases from other species.

CBG contains several sequence elements that are highly conserved among many *family 1* β-glucosidases. Between residues 34 and 48 it carries the N-terminal signature sequence
F,X,(FYWM),(GSTA),X,(GSTA),X,(GSTA),(GSTA),
(FYN),X,E,X(GSTA) characteristic of *family 1* glycosyl hydrolases (Henrissat, 1991). Two of the five cysteine

residues found in CBG (Cys175 and Cys225) are also conserved in these homologous β-glucosidases, suggesting that they may be involved in forming important intramolecular disulfide bridges.

Other conserved sequence elements include the sequence -ENG- at residues 408-410 within the C-terminal signature, and the sequence -NEP- at residues 190-192. These sequence motifs are thought to be important for enzyme activity, and this region may be involved in binding of the pyranose ring during catalysis. The NEP motif of both *Bacillus* endo-β-1-4-glucanase and CBG is flanked by hydrophobic amino acids; next to the signal peptide, it is the most hydrophobic region of the CBG enzyme. The hydrolytic mechanism of the *family 1* β-glucosidases is considered to be general acid catalysis (Sinnott, 1990) with Glu and Asp residues in conserved motifs serving as active site nucleophile and acid catalyst. Evidence from inhibitor and site-directed mutagenesis studies suggests the Glu408 within the conserved ENG motif is the active site nucleophile (Withers et al., 1990; Trimbur et al, 1992). A conserved aspartate residue (Asp427) located 19 residues downstream from the ENG motif of CBG appears to be analogous to Asp374 of *Agrobacterium* β-glucosidase (cellobiase). This carboxylate side-chain may play the role of acid-base catalyst during hydrolysis of the glycosidic linkage (Trimbur et al., 1992).

EXAMPLE 3
Expression of CBG cDNA in *E. coli*

To express CBG protein in *E. coli*, the full-length coding region for the mature protein (i.e. excluding the signal peptide) was amplified using the 3'end clone (1A6) as the template with the N-terminal primer, NT1 (5' TAGCTAGCAGGCTGGACAGGAACAACTTC 3', Seq. I.D. No. 7) containing a 5' Nhe1 site, and a C-terminal primer, CT1 (5' CTCGAGACAAGCAGTCTAAATGCT 3', Seq. I.D. No. 8) containing a Xho1 site. The resulting 1.5kb DNA fragment was ligated into Bluescript II KS by T/A cloning as described above. The structure of the junctions of this construct was confirmed by sequencing and it was then inserted as a Nhe1/Xho1 fragment into expression vector pET21a (Novagen). Because, the Nhe1 site was used to introduce the cDNA into the pET vector, three non-CBG amino acids (Met, Ala, Ser) were added to the N-terminus of the expressed protein. To avoid the expression of the vector His-tag at the 3' end, the native stop codon of CBG was included. The expressed protein was thus identical in sequence to the mature CBG expressed *in planta*, except for the additional tripeptide at the N-terminus. Following transformation into *E.coli* strain DH5α and verification of the plasmid integrity by restriction digestion, it was introduced into the expression host BL21(DE3).

To express CBG, the bacteria were grown to log phase (A_{600}=0.6-0.9) followed by an additional 2-3 h incubation at 29-37°C in the presence or absence of 0.4-

1mM IPTG. The expressed CBG in the soluble protein fraction was purified by preparative Q-Sepharose chromatography followed by QMA-Memsep (Millipore) chromatography.

As noted above, the functional characteristic of the CBG enzyme is its ability to hydrolyze coniferin. This activity can be measured using the simple β-glucosidase assay described by Dharwardhana et al. (1995), conducted as follows: enzyme preparations (10-50 μl) and glucoside substrate (coniferin) (2mM final concentration) in 0.2M MES, pH 5.5 buffer in a final volume of 150 μl are incubated at 30°C for 30 min. The reaction is stopped by basification of the assay mixture with an equal volume of 0.5M CAPS buffer (Sigma Chemical Co., St. Louis, MO), pH 10.5 and the activity measured by determining the absorbance of the released aglycone. The activity of the enzyme can be measured not only against coniferin, but also against related glucosides including 4-NPG, 2-NPG, MUG and the synthetic coniferin analog VRA-G. For quantitative calculations, the following analysis wavelengths and ϵ values ($mM^{-1}x\ cm^{-1}$) were used: coniferyl alcohol, 325nm, ϵ = 7.0; sinapyl alcohol, 315nm, ϵ = 11.2; 2-nitrophenol, 420nm, ϵ = 4.55, 4-nitrophenol, 400nm, ϵ =19.3; 4-methyl umbelliferone, 360 nm, ϵ = 18.25; VRA-G, 490nm, ϵ = 38.6; salicyl alcohol, 295nm, ϵ = 3.3.

Soluble proteins and insoluble proteins (inclusion bodies) prepared from induced and uninduced bacterial cells were assayed for coniferin hydrolysis

activity by the method described above. Only the soluble protein fraction of induced cells displayed this activity. The activity in this fraction could be increased up to 2-fold by increasing the IPTG concentration from 0.4 - 1.0 mM, and by reducing the growing temperature from 37°C to 29°C. Activity staining of nondenaturing PAGE gels using the chromogenic coniferin analogue VRA-G revealed a β-glucosidase-active protein band in induced cell extracts. This protein was purified by anion exchange chromatography using coniferin as the substrate for monitoring β-glucosidase activity. The purified enzyme often migrated as a doublet on nondenaturing gels. Both protein bands in the doublet showed β-glucosidase activity, as assayed by hydrolysis of VRA-G. This could be due to partial degradation, alternate forms of folding, or the synthesis of a truncated protein at the 5' end where CBG has a prokaryotic ribosome binding Shine-Dalgarno sequence (GAAGGAG). The latter would result in the synthesis of a polypeptide that is truncated at the N-terminus, as opposed to the full-length polypeptide initiated by ribosome binding to the standard ribosome binding site in the vector. As shown in Table 1 below, the CBG expressed in *E. coli* and the enzyme purified from the pine xylem showed almost identical substrate specificities.

Tabl 1. Substrate specificity of coniferin β-glucosidase purified from pine xylem and *E.coli*-expressed CBG-cDNA. 100% activity represents 14pKat for native coniferin β-glucosidase and 22pKat for the recombinant enzyme.

	Relative activity	
Substrate	**Native CBG**	***E.coli* CBG**
coniferin	100	100
syringin	51	65
4-methyl umbelliferyl-β-glucoside	18	20
2-nitrophenyl-β-glucoside	51	50
4-nitrophenyl-β-glucoside	30	35

EXAMPLE 4
Preferred Method for Making the CBG cDNA

With the provision of the CBG cDNA sequence shown in Seq. I.D. No. 6, the polymerase chain reaction (PCR) may now be utilized in a preferred method for producing the CBG cDNA. PCR amplification of the CBG cDNA sequence may be accomplished either by direct PCR from an appropriate cDNA library or by Reverse-Transcription PCR (RT-PCR) using RNA extracted from plant cells as a template. Methods and conditions for both direct PCR and RT-PCR are known in the art and are described in Innis et al. (1990). Suitable plant cDNA libraries for direct PCR include the *Pinus contorta* library as described above. Other plant cDNA libraries

may be used in order to amplify orthologous cDNAs of other species; for example, the Arabidopsis cDNA library described by Newman et al. (1994) may be used to amplify the Arabidopsis ortholog.

The selection of PCR primers will be made according to the portions of the cDNA which are to be amplified. Primers may be chosen to amplify small segments of the cDNA or the entire cDNA molecule. Variations in amplification conditions may be required to accommodate primers of differing lengths; such considerations are well known in the art and are discussed in Innis et al. (1990), Sambrook et al. (1989), and Ausubel et al (1992). By way of example only, the entire CBG cDNA molecule as shown in Seq. I.D. No. 6 may be amplified using the following combination of primers:

5' GGATTTGGACCTGAAAATATCAAT 3' (Seq. I.D. No. 9)

5' CAATGTTCTTACCCTGCAGTTCCC 3' (Seq. I.D. No. 10)

The open reading frame portion of the cDNA may be amplified using the following primer pair:

5' ATGGAGGTGTCTGTGTTGATGTGGGTA 3' (Seq. I.D. No. 11)

5' AATGCTGCTGCTGCTTCTAATACTTCC 3' (Seq. I.D. No. 12)

These primers are illustrative only; it will be appreciated by one skilled in the art that many different primers may be derived from the provided cDNA sequence in order to amplify particular regions of this cDNA. Suitable amplification conditions include those described above for the original isolation of the CBG cDNA. As is well known in the art, amplification

conditions may need be varied in order to amplify orthologous genes where the sequence identity is not 100%; in such cases, the use of nested primers, as described above may be beneficial. Resequencing of PCR products obtained by these amplification procedures is recommended; this will facilitate confirmation of the CBG cDNA sequence and will also provide information on natural variation on this sequence in different ecotypes, cultivars and plant populations.

Oligonucleotides which are derived from the CBG cDNA sequence and which are suitable for use as PCR primers to amplify the CBG cDNA are encompassed within the scope of the present invention. Preferably, such oligonucleotide primers will comprise a sequence of 15-20 consecutive nucleotides of the CBG cDNA. To enhance amplification specificity, primers of 20-30 nucleotides or more in length may also be used.

EXAMPLE 5
Use of the CBG cDNA to Produce Plants with Modified Lignin Content

Once a gene (or cDNA) encoding a protein involved in the determination of a particular plant characteristic has been isolated, standard techniques may be used to express the cDNA in transgenic plants in order to modify that particular plant characteristic. The basic approach is to clone the cDNA into a transformation vector, such that it is operably linked to control sequences (e.g., a promoter) which direct

expression of the cDNA in plant cells. The transformation vector is then introduced into plant cells by one of a number of techniques (e.g., electroporation) and progeny plants containing the introduced cDNA are selected. Preferably all or part of the transformation vector will stably integrate into the genome of the plant cell. That part of the transformation vector which integrates into the plant cell and which contains the introduced cDNA and associated sequences for controlling expression (the introduced "transgene") may be referred to as the recombinant expression cassette.

Selection of progeny plants containing the introduced transgene may be made based upon the detection of an altered phenotype. Such a phenotype may result directly from the cDNA cloned into the transformation vector or may be manifested as enhanced resistance to a chemical agent (such as an antibiotic) as a result of the inclusion of a dominant selectable marker gene incorporated into the transformation vector.

The choice of (a) control sequences and (b) how the cDNA (or selected portions of the cDNA) are arranged in the transformation vector relative to the control sequences determine, in part, how the plant characteristic affected by the introduced cDNA is modified. For example, the control sequences may be tissue specific, such that the cDNA is only expressed in particular tissues of the plant (e.g., vascular systems) and so the affected characteristic will be modified only

in those tissues. The cDNA sequence may be arranged relative to the control sequence such that the cDNA transcript is expressed normally, or in an antisense orientation. Expression of an antisense RNA that is the reverse complement of the cloned cDNA will result in a reduction of the targeted gene product (the targeted gene product being the protein encoded by the plant gene from which the introduced cDNA was derived). Over-expression of the introduced cDNA, resulting from a plus-sense orientation of the cDNA relative to the control sequences in the vector, may lead to an increase in the level of the gene product, or may result in a reduction in the level of the gene product due to co-suppression (also termed "sense suppression") of that gene product.

Successful examples of the modification of plant characteristics by transformation with cloned cDNA sequences are replete in the technical and scientific literature. Selected examples, which serve to illustrate the level knowledge in this field of technology include:

U.S. Patent No. 5,451,514 to Boudet (modification of lignin synthesis using antisense RNA and co-suppression);

U.S. Patent No. 5,443,974 to Hitz (modification of saturated and unsaturated fatty acid levels using antisense RNA and co-suppression);

U.S. Patent No. 5,530,192 to Murase (modification of amino acid and fatty acid composition

using antisense RNA);

U.S. Patent No. 5,455,167 to Voelker (modification of medium chain fatty acids)

U.S. Patent No. 5,231,020 to Jorgensen (modification of flavonoids using co-suppression); and

U.S. Patent No. 5,583,021 to Dougherty (modification of virus resistance by expression of plus-sense RNA)

These examples include descriptions of transformation vector selection, transformation techniques and the construction of constructs designed to over-express the introduced cDNA, untranslatable RNA forms or antisense RNA. In light of the foregoing and the provision herein of the CBG cDNA, it is thus apparent that one of skill in the art will be able to introduce this cDNA, or derivative forms of the cDNA (e.g., antisense forms), into plants in order to produce plants having modified lignin content. Example 6 below provides an exemplary illustration of how an antisense form of the CBG cDNA may be introduced into conifers using ballistic transformation, in order to produce coniters having altered lignin content.

a. Plant Types

Lignins are found in all plant types, and thus DNA molecules according to the present invention (e.g., the CBG cDNA, homologs of the CBG cDNA and antisense forms) may be introduced into any plant type in order to modify the lignin composition of the plant. Thus, the

sequences of the present invention may be used to modify lignin composition in any higher plants including monocotyledonous plants such as lily, corn, rice, wheat and barley as well as dicotyledonous plants, such as tomato, potato, soy bean, cotton, tobacco, sunflower, safflower and brasicca. As noted above, the present invention is expected to be particularly useful in woody species such as species belonging to the genera *Picea*, *Pseudotsuga*, *Tsuga*, *Sequoia*, *Abies*, *Thuja*, *Libocedrus*, *Chamaecyparis* and *Laryx*. Pines are expected to be a particularly suitable choice for genetic modification by the methods disclosed herein, including lodgepole pine (*Pinus contorta*), the species from which the CBG cDNA was cloned.

b. Vector Construction, Choice of Promoters

A number of recombinant vectors suitable for stable transfection of plant cells or for the establishment of transgenic plants have been described including those described in Pouwels et al., (1987), Weissbach and Weissbach, (1989), and Gelvin et al., (1990). Typically, plant transformation vectors include one or more cloned plant genes (or cDNAs) under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable marker. Such plant transformation vectors typically also contain a promoter regulatory region (e.g., a regulatory region controlling inducible or constitutive, environmentally-or developmentally-regulated, or cell- or tissue-specific

expression), a transcription initiation start site, a ribosome binding site, an RNA processing signal, a transcription termination site, and/or a polyadenylation signal.

Examples of constitutive plant promoters which may be useful for expressing the CBG cDNA include: the cauliflower mosaic virus (CaMV) 35S promoter, which confers constitutive, high-level expression in most plant tissues (*see, e.g.*, Odel et al., 1985; Dekeyser et al., 1990; Terada and Shimamoto, 1990); the nopaline synthase promoter (An et al., 1988); and the octopine synthase promoter (Fromm et al., 1989).

A variety of plant gene promoters that are regulated in response to environmental, hormonal, chemical, and/or developmental signals, also can be used for expression of the CBG cDNA in plant cells, including promoters regulated by: (a) heat (Callis et al., 1988); (b) light (*e.g.*, the pea rbcS-3A promoter, Kuhlemeier et al., 1989, the maize rbcS promoter, Schaffner and Sheen, 1991, and the chlorophyll a/b-binding protein promoter); (c) hormones, such as abscisic acid (Marcotte et al., 1989); (d) wounding (*e.g.*, *wunI*, Siebertz et al., 1989); and (e) chemicals such as methyl jasminate or salicylic acid. It may also be advantageous to employ tissue-specific promoters, such as those described by Roshal et al., (1987), Schernthaner et al., (1988), and Bustos et al., (1989).

Plant transformation vectors may also include RNA processing signals, for example, introns, which may

be positioned upstream or downstream of the CBG cDNA sequence in the transgene. In addition, the expression vectors may also include additional regulatory sequences from the 3'-untranslated region of plant genes, *e.g.*, a 3' terminator region to increase mRNA stability of the mRNA, such as the PI-II terminator region of potato or the octopine or nopaline synthase 3' terminator regions.

Finally, as noted above, plant transformation vectors may also include dominant selectable marker genes to allow for the ready selection of transformants. Such genes include those encoding antibiotic resistance genes (*e.g.*, resistance to hygromycin, kanamycin, bleomycin, G418, streptomycin or spectinomycin) and herbicide resistance genes (*e.g.*, phosphinothricin acetyltransferase).

c. Arrangement of CBG cDNA in Vector

As noted above, the particular arrangement of the CBG cDNA in the transformation vector will be selected according to the expression of the cDNA desired.

Sense Expression

Where enhanced lignin synthesis is desired, the CBG cDNA may be operably linked to a constitutive high-level promoter such as the CaMV 35S promoter. As noted below, modification of lignin synthesis may also be achieved by introducing into a plant a transformation vector containing a variant form of the CBG cDNA, for example a form which varies from the exact nucleotide

sequence of the CBG cDNA, but which encodes a protein that retains the functional characteristic of the CBG protein, i.e. coniferin hydrolysis activity.

Sense Suppression

Constructs in which the CBG cDNA (or variants thereon) are over-expressed may also be used to obtain co-suppression of the endogenous CBG gene in the manner described in U.S. Patent No. 5,231,021 to Jorgensen. Such co-suppression (also termed sense suppression) does not require that the entire CBG cDNA be introduced into the plant cells, nor does it require that the introduced sequence be exactly identical to the CBG cDNA. However, as with antisense suppression, the suppressive efficiency will be enhanced as (1) the introduced sequence is lengthened and (2) the sequence similarity between the introduced sequence and the endogenous CBG geneis increased. Sense-suppression is believed to be modulated, in part, by the position on the plant genome into which the introduced sequence integrates.

Antisense Expression

In contrast, a reduction of lignin synthesis may be obtained by introducing antisense constructs based on the CBG cDNA sequence into plants. For antisense suppression, the CBG cDNA is arranged in reverse orientation relative to the promoter sequence in the transformation vector. The introduced sequence need not be the full length CBG cDNA, and need not be exactly homologous to the CBG cDNA. Generally, however, where

the introduced sequence is of shorter length, a higher degree of homology to the native CBG sequence will be needed for effective antisense suppression. Preferably, the introduced antisense sequence in the vector will be at least 30 nucleotides in length, and improved antisense suppression will typically be observed as the length of the antisense sequence increases. Preferably, the length of the antisense sequence in the vector will be greater than 100 nucleotides. Transcription of an antisense construct as described results in the production of RNA molecules that are the reverse complement of mRNA molecules transcribed from the endogenous CBG gene in the plant cell. Although the exact mechanism by which antisense RNA molecules interfere with gene expression has not been elucidated, it is believed that antisense RNA molecules bind to the endogenous mRNA molecules and thereby inhibit translation of the endogenous mRNA.

Suppression of endogenous CBG gene expression can also be achieved using ribozymes. Ribozymes are synthetic RNA molecules that possess highly specific endoribonuclease activity. The production and use of ribozymes are disclosed in U.S. Patent No. 4,987,071 to Cech and U.S. Patent No. 5,543,508 to Haselhoff. The inclusion of ribozyme sequences within antisense RNAs may be used to confer RNA cleaving activity on the antisense RNA, such that endogenous mRNA molecules that bind to the antisense RNA are cleaved, which in turn leads to an enhanced antisense inhibition of endogenous

gene expression.

Untranslatable RNA

Suppression of native gene expression may be achieved by transforming the plant with a sequence that is homologous to the target gene, but which is rendered untranslatable by a genetic modification such as the introduction of a premature stop codon. This approach is described in U.S. Patent No. 5,583,021. The introduced CBG sequence is preferably 50-100 nucleotides in length, although longer sequences, such as 100-250 nucleotides are preferred. The introduced sequence is engineered to encode an untranslatable RNA; the introduction of a premature stop codon early on in the coding region is a preferred way of achieving this. The sequence need not be perfectly homologous to the target CBG sequence, but at least 80%, and preferably 85% sequence homology will likely be more effective than lower homologies.

d. Transformation and Regeneration Techniques

Transformation and regeneration of both monocotyledonous and dicotyledonous plant cells are now routine, and the selection of the most appropriate transformation and regeneration techniques will be determined by the practitioner. The choice of method will vary with the type of plant to be transformed; those skilled in the art will recognize the suitability of particular methods for given plant types. Suitable

methods may include, but are not limited to: electroporation of plant protoplasts; liposome-mediated transformation; polyethylene mediated transformation; transformation using viruses; micro-injection of plant cells; micro-projectile bombardment of plant cells; vacuum infiltration; and *Agrobacterium tumeficiens* (AT) mediated transformation. Typical procedures for transforming and regenerating plants are described in the patent documents listed at the beginning of this section. In addition, methods for transforming woody species are described in Ellis et al. (1993), Ellis et al. (1996), U.S. Patent No. 5,122,466, "Ballistic Transformation of Conifer" and U.S. Patent No. 4,795,855, "Transformation and Foreign Gene Expression with Woody Species".

e. Selection of Transformed Plants

Following transformation and regeneration of plants with the transformation vector, transformed plants are preferably selected using a dominant selectable marker incorporated into the transformation vector. Typically, such a marker will confer antibiotic resistance on the seedlings of transformed plants, and selection of transformants can be accomplished by exposing the seedlings to appropriate concentrations of the antibiotic.

After transformed plants are selected and grown to maturity, they can be assayed to determine whether coniferin β-glucosidase synthesis has been

altered as a result of the introduced transgene. This can be done in several ways, including by extracting and quantifying the enzyme activity as described in Example 6. In addition, lignification may be determined histochemically, and lignin content may be quantified, as described in Example 6. Also, antisense or sense suppression of the endogenous CBG gene may be detected by analyzing mRNA expression on Northern blots.

EXAMPLE 6
Introduction Of Antisense CBG cDNA Sequence Into White Spruce (*Picea Glauca*)

By way of example, the following methodology may be used to produce white spruce trees having an altered lignin content. The CBG cDNA is operably linked, but in reverse orientation, to the enhanced cauliflower mosaic virus (CaMV) 35S promoter in place of the BT gene in plasmid pTVBT41100 (Ellis et al., 1993). (Many other plants tranaformation vectors have been described and would be suitable for introducing CBG-based constructs into plants. Vector pBACGGUS shown in Fig. 4 is one such alternative vector that may be used). Somatic embryos of *Picea glauca* are differentiated from embryogenic white spruce callus line and cultured as described by Ellis et al. (1993). Plasmid DNA is adhered to 1-3μM gold particles (0.5μg DNA / mg gold) by calcium chloride and spermidine precipitation. Gold particles containing the DNA are then loaded on to carrier sheets at a rate of 0.05mg/cm² and these particles are then introduced into somatic embryos as described by Ellis et

al. (1991). Transformed embryos are selected using kanamycin. Regeneration of transgenic plants (via the production of embryogenic callus) is achieved using the culture conditions described by Ellis et al. (1993).

In order to determine coniferin β-glucosidase activity in the transgenic plants, the enzyme is extracted as described in Example 1 above, and the activity is assayed using the β-glucosidase assay described in Example 3 above. Plants transformed with the same vector without the CBG cDNA insert should preferably be used as controls. In situ localization of the enzyme activity can be determined using VRA-G as described by Dharmawardhana et al. (1995). Lignin in the stem sections is detected histochemically by Basic Fuchsin-induced fluorescence and imaging on a confocal laser scanning microscope as described by Dharmawardhana et al. (1992). In order to determine the effect of introducing the antisense construct into the plant on lignin content, standard methods are used to quantify lignin in the transformed plant (and control plants). Standard methods of quantifying lignin include the thioglycolic acid procedure as described by Whitmore (1978) and the acetyl bromide procedure as described by Liyama and Wallis (1990).

EXAMPLE 7
Production of Sequence Variants

As noted above, modification of lignin synthesis in plant cells can be achieved by transforming

plants with the CBG cDNA, antisense constructs based on the CBG cDNA or other variants on the CBG cDNA sequence. With the provision of the CBG cDNA sequence herein, the creation of variants on the CBG cDNA sequence by standard mutagenesis techniques is now enabled.

Variant DNA molecules include those created by standard DNA mutagenesis techniques, for example, M13 primer mutagenesis. Details of these techniques are provided in Sambrook et al. (1989), Ch. 15. By the use of such techniques, variants may be created which differ in minor ways from the CBG cDNA. DNA molecules and nucleotide sequences which are derivatives of those specifically disclosed herein and which differ from those disclosed by the deletion, addition or substitution of nucleotides while still encoding a protein which possesses the functional characteristic of the CBG protein (i.e., the ability to hydrolyze coniferin) are comprehended by this invention. DNA molecules and nucleotide sequences which are derived from the CBG cDNA include DNA sequences which hybridize under moderately stringent conditions to the DNA sequences disclosed, or fragments thereof.

Hybridization conditions resulting in particular degrees of stringency will vary depending upon the nature of the hybridization method of choice and the composition and length of the hybridizing DNA used. Generally, the temperature of hybridization and the ionic strength (especially the Na^+ concentration) of the hybridization buffer will determine the stringency

of hybridization. Calculations regarding hybridization conditions required for attaining particular degrees of stringency are discussed by Sambrook et al. (1989), chapters 9 and 11. By way of illustration only, a hybridization experiment may be performed by hybridization of a DNA molecule (for example, a variation of the CBG cDNA sequence) to a target DNA molecule (for example, the native CBG cDNA sequence) which has been electrophoresed in an agarose gel and transferred to a nitrocellulose membrane by Southern blotting (Southern, 1975), a technique well known in the art and described in (Sambrook et al., 1989). Hybridization with a target probe labeled with [^{32}P]-dCTP is generally carried out in a solution of high ionic strength such as 6xSSC at a temperature that is 20-25°C below the melting temperature, T_m, described below. For such Southern hybridization experiments where the target DNA molecule on the Southern blot contains 10 ng of DNA or more, hybridization is typically carried out for 6-8 hours using 1-2 ng/ml radiolabeled probe (of specific activity equal to 10^9 CPM/μg or greater). Following hybridization, the nitrocellulose filter is washed to remove background hybridization. The washing conditions should be as stringent as possible to remove background hybridization but to retain a specific hybridization signal. The term T_m represents the temperature above which, under the prevailing ionic conditions, the radiolabeled probe molecule will not hybridize to its target DNA molecule. The T_m of such a hybrid molecule

may be estimated from the following equation (Bolton and McCarthy, 1962):

T_m = 81.5°C - 16.6($\log_{10}$[Na^+]) + 0.41(%G+C) - 0.63(% formamide) - (600/l)

Where l = the length of the hybrid in base pairs. This equation is valid for concentrations of Na^+ in the range of 0.01 M to 0.4 M, and it is less accurate for calculations of T_m in solutions of higher [Na^+]. The equation is also primarily valid for DNAs whose G+C content is in the range of 30% to 75%, and it applies to hybrids greater than 100 nucleotides in length (the behavior of oligonucleotide probes is described in detail in Ch. 11 of Sambrook et al., 1989).

Thus, by way of example, for a 150 base pair DNA probe derived from the first 150 base pairs of the open reading frame of the CBG cDNA (with a hypothetical %GC = 45%), a calculation of hybridization conditions required to give particular stringencies may be made as follows:

For this example, it is assumed that the filter will be washed in 0.3x SSC solution following hybridization, thereby

[Na^+] = 0.045M

%GC = 45%

Formamide concentration = 0

l = 150 base pairs

$$T_m = 81.5 - 16(\log_{10}[Na^+]) + (0.41 \times 45) - \left(\frac{600}{150}\right)$$

and so T_m = 74.4°C.

The T_m of double-stranded DNA decreases by

1-1.5°C with every 1% decrease in homology (Bonner et al., 1973). Therefore, for this given example, washing the filter in 0.3x SSC at 59.4-64.4°C will produce a stringency of hybridization equivalent to 90%; that is, DNA molecules with more than 10% sequence variation relative to the target CBG cDNA will not hybridize. Alternatively, washing the hybridized filter in 0.3 xSSC at a temperature of 65.4-68.4°C will yield a hybridization stringency of 94%; that is, DNA molecules with more than 6% sequence variation relative to the target CBG cDNA molecule will not hybridize. The above example is given entirely by way of theoretical illustration. One skilled in the art will appreciate that other hybridization techniques may be utilized and that variations in experimental conditions will necessitate alternative calculations for stringency.

As used herein, moderate stringency conditions are those under which DNA molecules with more than 25% sequence variation (also termed "mismatch") will not hybridize. As noted above, the invention encompasses DNA molecules which hybridize under moderately stringent conditions to the CBG cDNA sequence. More preferably, such DNA molecules will hybridize under stringent conditions, which are conditions under which DNA molecules with more than 15% mismatch will not hybridize. More preferably still, such DNA molecules will hybridize under highly stringent conditions, i.e., those under which DNA sequences with more than 10% mismatch will not hybridize. Finally, in the most

preferred embodiment, these DNA molecules will hybridize to the CBG cDNA under extremely stringent conditions, that is, conditions under which DNA sequences with more than 6% mismatch will not hybridize.

The degeneracy of the genetic code further widens the scope of the present invention as it enables major variations in the nucleotide sequence of a DNA molecule while maintaining the amino acid sequence of the encoded protein. For example, the 23rd amino acid residue of the CBG protein is alanine. This is encoded in the CBG cDNA by the nucleotide codon triplet GCT. Because of the degeneracy of the genetic code, three other nucleotide codon triplets--GCA, GCC and GCG--also code for alanine. Thus, the nucleotide sequence of the CBG cDNA could be changed at this position to any of these three codons without affecting the amino acid composition of the encoded protein or the characteristics of the protein. The genetic code and variations in nucleotide codons for particular amino acids is presented in Tables 2 and 3. Based upon the degeneracy of the genetic code, variant DNA molecules may be derived from the cDNA molecules disclosed herein using standard DNA mutagenesis techniques as described above, or by synthesis of DNA sequences. Thus, this invention also encompasses DNA sequences which encode the CBG protein but which vary from the CBG cDNA sequence by virtue of the degeneracy of the genetic code.

TABLE 2
The Genetic Cod

First Position (5' end)	Second Position T	Second Position C	Second Position A	Second Position G	Third Position (3' end)
T	Phe	Ser	Tyr	Cys	T
	Phe	Ser	Tyr	Cys	C
	Leu	Ser	Stop (och)	Stop	A
	Leu	Ser	Stop (amb)	Trp	G
C	Leu	Pro	His	Arg	T
	Leu	Pro	His	Arg	C
	Leu	Pro	Gln	Arg	A
	Leu	Pro	Gln	Arg	G
A	Ile	Thr	Asn	Ser	T
	Ile	Thr	Asn	Ser	C
	Ile	Thr	Lys	Arg	A
	Met	Thr	Lys	Arg	G
G	Val	Ala	Asp	Gly	T
	Val	Ala	Asp	Gly	C
	Val	Ala	Glu	Gly	A
	Val (Met)	Ala	Glu	Gly	G

"Stop (och)" stands for the ocre termination triplet, and "Stop (amb)" for the amber. ATG is the most common initiator codon; GTG usually codes for valine, but it can also code for methionine to initiate an mRNA chain.

TABLE 3
The Degeneracy of the Genetic Code

Number of Synonymous Codons	Amino Acid	Total Number of Codons
6	Leu, Ser, Arg	18
4	Gly, Pro, Ala, Val, Thr	20
3	Ile	3
2	Phe, Tyr, Cys, His, Gln, Glu, Asn, Asp, Lys	18
1	Met, Trp	2
Total number of codons for amino acids		61
Number of codons for termination		3
Total number of codons in genetic code		64

One skilled in the art will recognize that DNA mutagenesis techniques may be used not only to produce variant DNA molecules, but will also facilitate the production of proteins which differ in certain structural aspects from the CBG protein, yet which are clearly derivative of the CBG protein and which maintain the essential characteristics of the CBG protein. Newly derived proteins may also be selected in order to obtain variations on the characteristic of the CBG protein, as will be more fully described below. Such derivatives include those with variations in amino acid sequence

including minor deletions, additions and substitutions.

While the site for introducing an amino acid sequence variation is predetermined, the mutation *per se* need not be predetermined. For example, in order to optimize the performance of a mutation at a given site, random mutagenesis may be conducted at the target codon or region and the expressed protein variants screened for the optimal combination of desired activity. Techniques for making substitution mutations at predetermined sites in DNA having a known sequence as described above are well known.

Amino acid substitutions are typically of single residues; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to 30 residues. Substitutions, deletions, insertions or any combination thereof may be combined to arrive at a final construct. Obviously, the mutations that are made in the DNA encoding the protein must not place the sequence out of reading frame and preferably will not create complementary regions that could produce secondary mRNA structure.

Substitutional variants are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the following Table 4 when it is desired to finely modulate the characteristics of the protein. Table 4 shows amino acids which may be substituted for an original amino acid in a protein and which are regarded

as conservative substitutions.

TABLE 4

Original Residue	Conservative Substitutions
Ala	ser
Arg	lys
Asn	gln, his
Asp	glu
Cys	ser
Gln	asn
Glu	asp
Gly	pro
His	asn; gln
Ile	leu. val
Leu	ile; val
Lys	arg; gln; glu
Met	leu; ile
Phe	met; leu; tyr
Ser	thr
Thr	ser
Trp	tyr
Tyr	trp; phe
Val	ile; leu

Substantial changes in enzymatic function or other features are made by selecting substitutions that are less conservative than those in Table 4, i.e., selecting residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. The

substitutions which in general are expected to produce the greatest changes in protein properties will be those in which (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histadyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

The effects of these amino acid substitutions or deletions or additions may be assessed for derivatives of the CBG protein by analyzing the ability of the derivative proteins to hydrolyze coniferin by the assay described herein.

REFERENCES

Altschul et al. (1990). *J.Mol. Biol.* **215**: 403-410
An et al. (1988) *Plant Physiol.* **88**:547
Ausubel et al. (1987). In *Current Protocols in Molecular Biology*, Greene Publishing Associates and Wiley-Intersciences.
Baird et al. (1990) *Biochem. Biophy. Res. Comm.* **169**: 1035-1039
Bolton and McCarthy (1962) *Proc. Natl. Acad. Sci. USA* **48**:1390.
Bonner et al. (1973). *J. Mol. Biol.* **81**:123.
Brzoboha et al. (1993) *Science* **262**: 1051
Bustos et al. (1989) *Plant Cell* **1**:839
Callis et al. (1988) *Plant Physiol.* **88**:965
Campbell & Ellis (1991) *Planta* **180**: 409-417
Castle et al. (1992) *J. Bacteriol.* **174**, 1478-1486
Christou (1996) *Trends Plant Sci.* **1**, 423-431
Dekeyser et al. (1990) *Plant Cell* **2**:591
Dharmawardhana et al. (1995) *Plant Physiol.* **107**: 331-339
Dharmawardhana et al. (1992) *Can. J. Bot.* **70**: 2238-2244
Elkind et al. (1990) *Proc Nat Acad Sci USA* **87**: 9057-9061
Ellis et al. (1993) Bio/Technology **11**: 84-89
Ellis et al. (1996) *In* Somatic Cell Genetics and Molecular Genetics of Trees, Boerjan and Ahuja (Eds.), Kluwer Academic Publishers, The Netherlands
Ellis et al. (1991) *Plant. Mol. Biol.* **17**: 19-27
Falk et al. (1992) *Plant Sci.* **83**, 181-186
Fan and Conn (1985) *Arch Biochem Biophys* **243**: 361-373
Freudenberg & Harkin (1963) *Phytochemistry* **2**: 189-193
Fromm et al. (1989) *Plant Cell* **1**:977
Fukushima & Terashima (1990) *J. Wood Chem. Technol.* **10**, 413-433
Fukushima & Terashima (1991) *Holzforschung.* **45**, 87-89
Gelvin et al. (1990) *Plant Molecular Biology Manual*, Kluwer Academic Publishers.
Harkin and Obst (1973) *Science* **180**: 296-297
Henrissat (1991) *Biochem. J.*, **280**: 309-316
Holton & Graham (1991) *Nucl. Acids Res.* **191**: 1156-1158
Hosel et al. (1982) *Plant Cell Orgn Cult* **1**:137-148
Hosel et al. (1987) *Arch Biochem Biophys* **252**: 152-162
Hosel and Todenhag n (1980) *Phytochemistry* **19**:1349-1353
Hrazdina & Wagn r (1985) *Arch. Biochem. Biophys.* **237**:

88-100.
Hrazdina & Jensen (1992) *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **43**: 241-267
Hughes & Dunn (1982) *Plant Mol. Biol.* **1**: 169-181.
Innis et al. (1990). *PCR Protocols, A Guide to Methods and Applications*, Innis et al. (eds.), Academic Press, Inc., San Diego, California.
Jefferson et al. (1981) *EMBO Journal* **6**: 3901
Joshi (1987) *Nuc. Acids Res.* **15**: 9627-9640
Klein et al. (1985) *Biochem. et Biophys. Acta.* **815**: 468-476
Kuhlemeier et al. (1989) *Plant Cell* **1**: 471
Laemmli (1970) *Nature* **277**: 680-6851
Lawton et al. (1990) *Plant Cell Rep.* **8**: 561-564
Leah et al. (1995) *J. Biol. Chem.* **270**: 15789-15796
Lee and Douglas (1994) *Plant Physiol suppl* **105**: 37
Leinhos & Savidge (1993) *Can J. For. Res.* **23**: 343-348
Leinhos et al. (1994) *Phytochemistry* **37**: 311-315
Lewinsohn et al. (1994) *Plant Mol. Biol. Rep.* **12**: 20-25
Li et al. (1992) *Plant Physiol* **100**: 282-290
Liyama & Wallis (1990) *J. Sci. Food Agriculture* **51**: 145-161
Marcinowski and Grisebach (1978) *Eur J Biochem* **87**: 37-44
Marcotte et al. (1989) *Plant Cell* **1**:969
Muckerheide et al. (1987) *J. Immunol* **138**: 833-837
Newman et al. (1994) *Plant Physiol.* **106**: 1241-1255
Odel et al. (1985) *Nature* 313:810.
Ohl et al. (1990) *Plant Cell* **2**, 837-848
Oxtoby et al. (1991) *Plant Mol. Biol.* **17**, 209-219.
Pearson and Lipman (1988) *Proc Nat Acad Sci USA* **85**: 2444-2448
Pickett-Heaps (1968) *Protoplasma* **65**: 181-190
Pouwels et al. (1987) *Cloning Vectors: A Laboratory Manual*, 1985, supp.
Roshal et al. (1987) *EMBO J.* **6**:1155.
Sambrook et al. (1989) Molecular cloning: A laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York
Savidge (1988) *Can. J. Bot.* **66**: 2099-2012
Savidge (1989) *Can J Bot* **67**: 2663-2668
Schernthaner et al. (1988) *EMBO J.* **7**:1249.
Schuch (1993) *Phytochem Soc North Am* **33 (1)**: 19
Schaffner and Sheen (1991) *Plant Cell* **3**:997
Siebertz et al. (1989) *Plant Cell* **1**:961

Simos t al. (1994) *Biochimica et Biophysica Acta* **1199**:52-58
Sinnott (1990) *Chem. Rev.* **90**: 1171-1202
Southern (1975) *J. Mol. Biol.* **98**:503.
Takabe et al. (1989) in *Plant Cell Wall Polymers, Biogenesis & Biodegradation*, Lewis N.G. & Paice M.G. (eds.) ACS Symp. Ser. 399. Amer. Chem. Soc. Washington DC pp. 47-66
Terada and Shimamoto (1990) *Mol. Gen. Genet.* **220**: 389
Terashima et al. (1986) *J. Wood Sci. Technol.* **6**: 495-504
Terashima & Fukushima (1988) *Wood Sci. Technol.* **22**, 259-270
Terazawa et al. (1984) *Mokuzai Gakkaishi* **30**: 322-328
Trimbur et al.(1992) *J. Biol. Chem.* **267**: 10248-10251
van-Uden et al. (1991) *Planta* **183**: 25-30
Varghese et al. (1994) *Proc. Natl. Acad. Sci. USA* **91**: 2785-2789
von Heijne (1986) *Nucl. Acids Res.* **14**: 4683-4690
Wagner et al. (1987) *Proc. Natl. Acad. Sci. USA* . **84**: 2097-2100
Weissbach and Weissbach (1989) *Methods for Plant Molecular Biology*, Academic Press.
Whetten & Sederoff (1995) *Plant Cell* **7**: 1001-1013
Whitmore (1978) *Phytochemistry* **17**: 412-425
Withers et al. (1990) *J. Amer. Chem. Soc.* **112**: 5887-5889
Wu et al. (1995) *Plant J.* **8**: 323-329
Xue et al. (1992) *Plant Mol. Biol.* **18**: 387-398
Yeoh and Woo (1992) *Phytochemistry* **31**: 2263-2265

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANTS: Carlson et al.

(ii) TITLE OF INVENTION: Coniferin Beta Glucosidase cDNA for Modifying Lignin Composition in Plants

(iii) NUMBER OF SEQUENCES: 12

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Klarquist Sparkman Campbell Leigh & Whinston, LLP

(B) STREET: One World Trade Center, Suite 1600, 121 S.W. Salmon Street

(C) CITY: Portland

(D) STATE: OR

(E) COUNTRY: USA

(F) ZIP: 97204-2988

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE:

(B) COMPUTER:

(C) OPERATING SYSTEM:

(D) SOFTWARE:

(vi) CURRENT APPLICATION DATA: Filed herewith

(A) APPLICATION NUMBER:

(B) FILING DATE:

(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER:

(B) FILING DATE:

(viii) ATTORNEY/AGENT INFORMATION

(A) NAME: David J Kapp, Ph.D.

(B) REGISTRATION NUMBER: P41,401

(C) REFERENCE/DOCKET NUMBER: 5493-46926/DJE

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (503) 226-7391

(B) TELEFAX: (503) 228-9446

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

GCTCTAGAGC GAYMGIAAYA AYTTTCC 27

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 27

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

GCTCTAGAGC GAYMGIAAYA AYTTCCC 27

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 25

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GAYMGIAAYA AYTTCCCIWS IGWTT 25

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 22

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

CACATATCTG TGATATTGGT CG 22

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

CCATCTTCTC GGACTGCTC 19

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 1909

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

```
ggatttggac ctgaaaatat caatttcaaa gcaattccag agggataacg          50
tgggatcctt accattacca acaacccacc attccgccct gccgacctca         100
ggcatatttt gattctattt aaccattaat tcatctgggc agttgtgatt         150
ctgtataatt cgatcgctcc gttttagcag                              180

ac atg gag gtg tct gtg ttg atg tgg gta ctg ctc ttc tat tcc     224
   Met Glu Val Ser Val Leu Met Trp Val Leu Leu Phe Tyr Ser
   1               5                  10

tta tta ggt ttt caa gtg acg aca gct agg ctg gac agg aac aac     269
Leu Leu Gly Phe Gln Val Thr Thr Ala Arg Leu Asp Arg Asn Asn
15                  20                  25

ttc ccc tca gat ttc atg ttc ggc aca gcc tct tca gcg tat cag     314
Phe Pro Ser Asp Phe Met Phe Gly Thr Ala Ser Ser Ala Tyr Gln
30                  35                  40

tat gaa gga gca gtc cga gaa gat ggc aag ggt cct agc aca tgg     359
Tyr Glu Gly Ala Val Arg Glu Asp Gly Lys Gly Pro Ser Thr Trp
45                  50                  55

gac gcc tta aca cat atg cct ggt aga ata aaa gat agc agc aat     404
Asp Ala Leu Thr His Met Pro Gly Arg Ile Lys Asp Ser Ser Asn
60                  65                  70
```

```
gga gac gtg gca gtc gac caa tat cac aga tat atg gaa gat atc      449
Gly Asp Val Ala Val Asp Gln Tyr His Arg Tyr Met Glu Asp Ile
75                  80                  85

gag ctt atg gct tca ctt gga cta gat gcc tat aga ttc tcc ata      494
Glu Ile Met Ala Ser Leu Gly Leu Asp Ala Tyr Arg Phe Ser Ile
90                  95                  100

tcc tgg tct cga atc ctt cca gaa gga aga ggt gaa att aac atg      539
Ser Trp Ser Arg Ile Leu Pro Glu Gly Arg Gly Glu Ile Asn Met
105                 110                 115

gct ggg att gaa tat tac aat aat ctg att gac gct ctt ctg caa      584
Ala Gly Ile Glu Tyr Tyr Asn Asn Leu Ile Asp Ala Leu Leu Gln
120                 125                 130

aat ggg atc cag ccg ttc gtg aca ttg ttc cat ttc gat ctt ccc      629
Asn Gly Ile Gln Pro Phe Val Thr Leu Phe His Phe Asp Leu Pro
135                 140                 145

aaa gca ctt gaa gac tcc tat ggg gga tgg ctg agt cct caa ata      674
Lys Ala Leu Glu Asp Ser Tyr Gly Gly Trp Leu Ser Pro Gln Ile
150                 155                 160

att aac gac ttc gaa gcc tat gca gag att tgc ttc cgg gca ttc      719
Ile Asn Asp Phe Glu Ala Tyr Ala Glu Ile Cys Phe Arg Ala Phe
165                 170                 175

ggt gac cgt gtc aaa tat tgg gcg aca gtg aac gag cca aat ctg      764
Gly Asp Arg Val Lys Tyr Trp Ala Thr Val Asn Glu Pro Asn Leu
180                 185                 190

ttt gtg ccg ttg gga tac acc gtc gga ata ttt cca ccg acg agg      809
Phe Val Pro Leu Gly Tyr Thr Val Gly Ile Phe Pro Pro Thr Arg
195                 200                 205

tgt gct gcc cct cac gcc aat cct ttg tgc atg aca ggg aat tgc      854
Cys Ala Ala Pro His Ala Asn Pro Leu Cys Met Thr Gly Asn Cys
210                 215                 220

tcg tca gca gag cca tat cta gct gca cat cac gtt ttg ctc gcc      899
Ser Ser Ala Glu Pro Tyr Leu Ala Ala His His Val Leu Leu Ala
225                 230                 235

cac gca tct gca gtg gag aaa tat agg gag aaa tat cag aaa att      944
His Ala Ser Ala Val Glu Lys Tyr Arg Glu Lys Tyr Gln Lys Ile
240                 245                 250

caa gga gga tct ata ggg tta gtt ata agc gcg cca tgg tac gaa      989
Gln Gly Gly Ser Ile Gly Leu Val Ile Ser Ala Pro Trp Tyr Glu
255                 260                 265
```

```
ccc ttg gaa aat tct cca gaa gag aga tca gct gtt gat aga att      1034
Pro Leu Glu Asn Ser Pro Glu Glu Arg Ser Ala Val Asp Arg Ile
270                 275                 280

tta tcc ttc aat ctc cga tgg ttt ttg gat cca att gtt ttt gga      1079
Leu Ser Phe Asn Leu Arg Trp Phe Leu Asp Pro Ile Val Phe Gly
285                 290                 295

gat tat cca caa gaa atg cgt gaa aga tta gga tcg cgc tta ccc      1124
Asp Tyr Pro Gln Glu Met Arg Glu Arg Leu Gly Ser Arg Leu Pro
300                 305                 310

tcc ata tcc tcg gaa cta tct gcg aaa ctt cgg gga tcg ttc gac      1169
Ser Ile Ser Ser Glu Leu Ser Ala Lys Leu Arg Gly Ser Phe Asp
315                 320                 325

tat atg ggt att aat cac tat aca acc tta tat gca aca agc act      1214
Tyr Met Gly Ile Asn His Tyr Thr Thr Leu Tyr Ala Thr Ser Thr
330                 335                 340

cct ccc ctt tcc ccc gac cac acg caa tat cta tat cca gac tct      1259
Pro Pro Leu Ser Pro Asp His Thr Gln Tyr Leu Tyr Pro Asp Ser
345                 350                 355

agg gtt tat ctg act gga gag cgc cac gga gtc tcc atc gga gaa      1304
Arg Val Tyr Leu Thr Gly Glu Arg His Gly Val Ser Ile Gly Glu
360                 365                 370

cgg aca ggg atg gac ggt ttg ttt gtg gta cct cat gga att caa      1349
Arg Thr Gly Met Asp Gly Leu Phe Val Val Pro His Gly Ile Gln
375                 380                 385

aaa ata gtg gag tat gta aaa gaa ttc tat gac aac ccg act att      1394
Lys Ile Val Glu Tyr Val Lys Glu Phe Tyr Asp Asn Pro Thr Ile
390                 395                 400

att atc gca gag aac ggt tat cca gag tct gag gaa tcc tcg tcg      1439
Ile Ile Ala Glu Asn Gly Tyr Pro Glu Ser Glu Glu Ser Ser Ser
405                 410                 415

act ctg caa gaa aat cta aac gat gtg agg aga ata agg ttt cat      1484
Thr Leu Gln Glu Asn Leu Asn Asp Val Arg Arg Ile Arg Phe His
420                 425                 430

gga gat tgt ttg agt tat ctc agt gca gca atc aaa aat ggc tca      1529
Gly Asp Cys Leu Ser Tyr Leu Ser Ala Ala Ile Lys Asn Gly Ser
435                 440                 445

gat gtt cga ggg tac ttt gtg tgg tca ctt ctg gat aat ttt gag      1574
Asp Val Arg Gly Tyr Phe Val Trp Ser Leu Leu Asp Asn Phe Glu
450                 455                 460
```

tgg gca ttt ggg tat acc att aga ttt ggt ctt tat cac gtg gat 1619
Trp Ala Phe Gly Tyr Thr Ile Arg Phe Gly Leu Tyr His Val Asp
465 470 475

ttc att tct gat caa aag aga tat ccc aag ctc tcg gct caa tgg 1664
Phe Ile Ser Asp Gln Lys Arg Tyr Phe Lys Leu Ser Ala Gln Trp
480 485 490

ttc aga caa ttt ctt cag cac gac gat cag gga agt att aga agc 1709
Phe Arg Gln Phe Leu Gln His Asp Asp Gln Gly Ser Ile Arg Ser
495 500 505

agc agc agc att tag actgcgttgt ctatttgcta atcaaagcgc 1754
Ser Ser Ser Ile
510

acacattcct gcaactctac ccaaaatcct gcaagcaaat atgttgtgtt 1804
cggatctatc caccgtgaga cacattacaa agaaatcatc aatctattcc 1854
aaaatgcaga aaaccccatt cagatgttct agggaactgc agggtaagaa 1904
cattg 1909

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

TAGCTAGCAG GCTGGACAGG AACAACTTC 29

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

CTCGAGACAA GCAGTCTAAA TGCT 24

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 24
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:
GGATTTGGAC CTGAAAATAT CAAT 24

(2) INFORMATION FOR SEQ ID NO: 10:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:
CAATGTTCTT ACCCTGCAGT TCCC 24

(2) INFORMATION FOR SEQ ID NO: 11:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:
ATGGAGGTGT CTGTGTTGAT GTGGGTA 27

(2) INFORMATION FOR SEQ ID NO: 12:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:
AATGCTGCTG CTGCTTCTAA TACTTCC 27

328434

CLAIMS

We claim:

1. An isolated nucleic acid molecule comprising at least 15 consecutive nucleotides of the sequence shown in Seq. I.D. No. 6 and encoding a coniferin β-glucosidase enzyme.

2. An isolated nucleic acid molecule according to claim 1 wherein the molecule comprises at least 20 consecutive nucleotides of the sequence shown in Seq. I.D. No. 6.

3. An isolated nucleic acid molecule according to claim 1 wherein the molecule comprises at least 30 consecutive nucleotides of the sequence shown in Seq. I.D. No. 6.

4. The isolated nucleic acid according to claim 1 wherein the nucleic acid molecule comprises the nucleotide sequence shown in Seq. I.D. No. 6.

5. An isolated nucleic acid molecule which encodes a coniferin β-glucosidase enzyme and which hybridizes under condition of at least moderate stringency to the nucleotide sequence shown in Seq. I.D. No. 6.

6. A coniferin β-glucosidase enzyme encoded by a

nucleic acid molecule according to claims 1-5.

7. A recombinant vector comprising a DNA sequence according to claims 1-5.

8. A transgenic plant transformed with a vector according to claim 7.

9. A transgenic plant according to claim 8 wherein the plant has an altered lignin content compared to an untransformed plant of the same species.

10. A transgenic plant according to claim 9 wherein the lignin content is reduced compared to an untransformed plant of the same species.

11. A transgenic plant according to claim 9 wherein the plant is a conifer.

12. A transgenic plant according to claim 9 wherein the plant is a *Pinus* species.

13. An isolated oligonucleotide which comprises at least 15 consecutive nucleotides of the sequence shown in Seq. I.D. No. 6 or its complementary strand.

14. An oligonucleotide according to claim 13 wherein the oligonucleotide comprises at least 30 consecutive nucleotides of the sequence shown in Seq.

I.D. No. 6 or its complementary strand.

15. An oligonucleotide according to claim 13 wherein the oligonucleotide comprises at least 100 consecutive nucleotides of the sequence shown in Seq. I.D. No. 6 or its complementary strand.

16. A recombinant vector comprising a DNA sequence according to claims 13-15.

17. A transgenic plant transformed with a vector according to claim 16.

18. A transgenic plant according to claim 17 wherein the plant has an altered lignin content compared to an untransformed plant of the same species.

19. A transgenic plant according to claim 18 wherein the lignin content is reduced compared to an untransformed plant of the same species.

20. A transgenic plant according to claim 19 wherein the plant is a conifer.

21. A transgenic plant according to claim 19 wherein the plant is a *Pinus* species.

22. A method of producing a plant with an altered lignin content relative to an untransformed plant

of that species, comprising introducing into the plant a recombinant vector that comprises a promoter operably linked to a nucleic acid which hybridizes under conditions of moderate stringency to the sequence shown in Seq. I.D. No. 6 and which encodes a coniferin β-glucosidase enzyme.

23. A transgenic plant produced according to the method of claim 22.

24. A transgenic plant comprising, integrated into its genome, a promoter operably linked to a nucleic acid which hybridizes under conditions of moderate stringency to the sequence shown in Seq. I.D. No 6 and which encodes a coniferin β-glucosidase enzyme.

25. A method of producing a plant with an altered lignin content relative to an untransformed plant of that species, comprising introducing into the plant a recombinant vector that comprises a promoter operably linked to an antisense nucleic acid which, when expressed in cells of the plant, inhibits the expression of a native coniferin β-glucosidase gene.

26. A transgenic plant produced according to the method of claim 25.

27. A transgenic plant comprising, integrated into its genome, a promoter operably linked to an

antisense nucleic acid which, when expressed in cells of the plant, inhibits the expression of a native coniferin β-glucosidase gene.

28. A method of producing a plant with an altered lignin content relative to an untransformed plant of that species, comprising introducing into the plant a nucleic acid molecule comprising a coding sequence operably linked to a promoter sequence, wherein the coding sequence encodes an untranslatable plus-sense transcript that shares at least 80% sequence similarity with a transcript of a native coniferin β-glucosidase gene.

29. A transgenic plant produced according to the method of claim 28.

30. A transgenic plant including, integrated into its genome, a nucleic acid molecule comprising a coding sequence operably linked to a promoter sequence, wherein the coding sequence encodes an untranslatable plus-sense transcript that shares at least 80% sequence similarity with a transcript of a native coniferin β-glucosidase gene.

31. An isolated nucleic acid which encodes a coniferin β-glucosidase.

32. The isolated nucleic acid according to claim

31 wherein the encoded coniferin β-glucosidase has an amino acid sequence as shown in Seq. I.D. No. 6.

33. A method of isolating a nucleotide sequence encoding a coniferin β-glucosidase enzyme, the method comprising hybridizing a nucleotide preparation with a DNA molecule comprising at least 15 consecutive nucleotides of the sequence set forth in Seq. I.D. No. 6.

34. An isolated nucleic acid molecule as defined in claim 1 substantially as herein described with reference to any example thereof and with or without reference to the accompanying drawings.

35. An isolated nucleic acid molecule as claimed in claim 5 substantially as herein described with reference to any example thereof and with or without reference to the accompanying drawings.

36. A coniferin β-glucosidase enzyme as claimed in claim 6 substantially as herein described with reference to any example thereof and with or without reference to the accompanying drawings.

37. A recombinant vector as claimed in claim 7 or claim 16 substantially as herein described with reference to any example thereof and with or without reference to the accompanying drawings.

38. A transgenic plant as defined in claim 8 or claim 17 substantially as herein described with reference to any example thereof and with or without reference to the accompanying drawings.

39. An isolated oligonucleotide as defined in claim 13 substantially as herein described with reference to any example thereof and with or without reference to the accompanying drawings.

40. A method as claimed in any one of claims 22, 25 and 28 of producing a plant with an altered lignin content relative to an untransformed plant of that species substantially as herein described with reference to any example thereof and with or without reference to the accompanying drawings.

41. A transgenic plant as claimed in any one of claims 23, 26, 27, 29 and 30 substantially as herein described with reference to any example thereof and with or without reference to the accompanying drawings.

42. An isolated nucleic acid as defined in claim 31 substantially as herein described with reference to any example thereof and with or without reference to the accompanying drawings.

43. A method as claimed in claim 33 of isolating a nucleotide sequence encoding a coniferin β-glucosidase enzyme substantially as herein described with reference to any example thereof and with or without reference to the accompanying drawings.

THE UNIVERSITY OF BRITISH COLUMBIA
by the authorised agents
A J Park & Son
Per

END OF CLAIMS

```
GGATTTGGACCTGAAAATATCAATTTCAAAGCAATTCCAGAGGGATAACGTGGGATCCTTACCATTACCAAC

AACCCACCATTCCGCCCTGCCGACCTCAGGCATATTTTGATTCTATTTAACCATTAATTCATCTGGGCAGTT

GTGATTCTGTATAATTCGATCGCTCCGTTTTAGCAGACATGGAGGTGTCTGTGTTGATGTGGGTACTGCTCT
                                                  M  E  V  S  V  L  M  W  V  L  L
TCTATTCCTTATTAGGTTTTCAAGTGACGACAGCTAGGCTGGACAGGAACAACTTCCCCTCAGATTTCATGT
F  Y  S  L  L  G  F  Q  V  T  T  A  R  L  D  R  N  N  F  P  S  D  F  M
TCGGCACAGCCTCTTCAGCGTATCAGTATGAAGGAGCAGTCCGAGAAGATGGCAAGGGTCCTAGCACATGGG
F  G  T  A  S  S  A  Y  Q  Y  E  G  A  V  R  E  D  G  K  G  P  S  T  W
ACGCCTTAACACATATGCCTGGTAGAATAAAAGATAGCAGCAATGGAGACGTGGCAGTCGACCAATATCACA
D  A  L  T  H  M  P  G  R  I  K  D  S  S  N  G  D  V  A  V  D  Q  Y  H
GATATATGGAAGATATCGAGCTTATGGCTTCACTTGGACTAGATGCCTATAGATTCTCCATATCCTGGTCTC
R  Y  M  E  D  I  E  L  M  A  S  L  G  L  D  A  Y  R  F  S  I  S  W  S
GAATCCTTCCAGAAGGAAGAGGTGAAATTAACATGGCTGGGATTGAATATTACAATAATCTGATTGACGCTC
R  I  L  P  E  G  R  G  E  I  N  M  A  G  I  E  Y  Y  N  N  L  I  D  A
TTCTGCAAAATGGGATCCAGCCGTTCGTGACATTGTTCCATTTCGATCTTCCCAAAGCACTTGAAGACTCCT
L  L  Q  N  G  I  Q  P  F  V  T  L  F  H  F  D  L  P  K  A  L  E  D  S
ATGGGGGATGGCTGAGTCCTCAAATAATTAACGACTTCGAAGCCTATGCAGAGATTTGCTTCCGGGCATTCG
Y  G  G  W  L  S  P  Q  I  I  N  D  F  E  A  Y  A  E  I  C  F  R  A  F
GTGACCGTGTCAAATATTGGGCGACAGTGAACGAGCCAAATCTGTTTGTGCCGTTGGGATACACCGTCGGAA
G  D  R  V  K  Y  W  A  T  V  N  E  P  N  L  F  V  P  L  G  Y  T  V  G
TATTTCCACCGACGAGGTGTGCTGCCCCTCACGCCAATCCTTTGTGCATGACAGGGAATTGCTCGTCAGCGG
I  F  P  P  T  R  C  A  A  P  H  A  N  P  L  C  M  T  G  N  C  S  S  A
AGCCATATCTAGCTGCACATCACGTTTTGCTCGCCCACGCATCTGCAGTGGAGAAATATAGGGAGAAATATC
E  P  Y  L  A  A  H  H  V  L  L  A  H  A  S  A  V  E  K  Y  R  E  K  Y
AGAAAATTCAAGGAGGATCTATAGGGTTAGTTATAAGCGCGCCATGGTACGAACCCTTGGAAAATTCTCCAG
Q  K  I  Q  G  G  S  I  G  L  V  I  S  A  P  W  Y  E  P  L  E  N  S  P
AAGAGAGATCAGCTGTTGATAGAATTTTATCCTTCAATCTCCGATGGTTTTTGGATCCAATTGTCTTCTGGAG
E  E  R  S  A  V  D  R  I  L  S  F  N  I  R  W  F  L  D  P  I  V  F  G
ATTATCCACAAGAAATGCGTGAAAGATTAGGATCGCGCTTACCCTCCATATCCTCGGAACTATCTGCGAAAC
D  Y  P  Q  E  M  R  E  R  L  G  S  R  L  P  S  I  S  S  E  L  S  A  K
TTCGGGGATCGTTCGACTATATGGGTATTAATCACTATACAACCTTATATGCAACAAGCACTCCTCCCCTTT
L  R  G  S  F  D  Y  M  G  I  N  H  Y  T  T  L  Y  A  T  S  T  P  P  L
CCCCCGACCACACGCAATATCTATATCCAGACTCTAGGGTTTATCTGACTGGAGAGCGCCACGGAGTCTCCA
S  P  D  H  T  Q  Y  L  Y  P  D  S  R  V  Y  L  T  G  E  R  H  G  V  S
TCGGAGAACGGACAGGGATGGACGGTTTGTTTGTGGTACCTCATGGAATTCAAAAAATAGTGGAGTATGTAA
I  G  E  R  T  G  M  D  G  L  F  V  V  P  H  G  I  Q  K  I  V  E  Y  V
AAGAATTCTATGACAACCCGACTATTATTATCGCAGAGAACGGTTATCCAGAGTCTGAGGAATCCTCGTCGA
K  E  F  Y  D  N  P  T  I  I  I  A  E  N  G  Y  P  E  S  E  E  S  S  S
CTCTGCAAGAAAATCTAAACGATGTGAGGAGAATAAGGTTTCATGGAGATTGTTTGATTTATCTCAGTGCAG
T  L  Q  E  N  L  N  D  V  R  R  I  R  F  H  G  D  C  L  S  Y  L  S  A
CAATCAAAAATGGCTCAGATGTTCGAGGGTACTTTGTGTGGTCACTTCTGGATAATTTTGAGTGGGCATTTG
A  I  K  N  G  S  D  V  R  G  Y  F  V  W  S  L  L  D  N  F  E  W  A  F
GGTATACCATTAGATTTGGTCTTTATCACGTGGATTTCATTCTGATCAAAAGAGATATCCCAAGCTCTCGG
G  Y  T  I  R  F  G  L  Y  H  V  D  F  I  S  D  Q  K  R  Y  P  K  L  S
CTCAATGGTTCAGACAATTTCTTCAGCACGACGATCAGGGAAGTATTAGAAGCAGCAGCAGCATTTAGACTG
A  Q  W  F  R  Q  F  L  Q  H  D  D  Q  G  S  I  R  S  S  S  S  I  -
CGTTGTCTATTTGCTAATCAAAGCGCACACATTCCTGCAACTCTACCCAAAATCCTGCAAGCAAATATGTTG

TGTTCGGATCTATCCACCGTGAGACACATTACAAAGAAATCATCAATCTATTCCAAAATGCAGAAAACCCCA

TTCAGATGTTCTAGGGAACTGCAGGGTAAGAACATTG
```

Figure 1

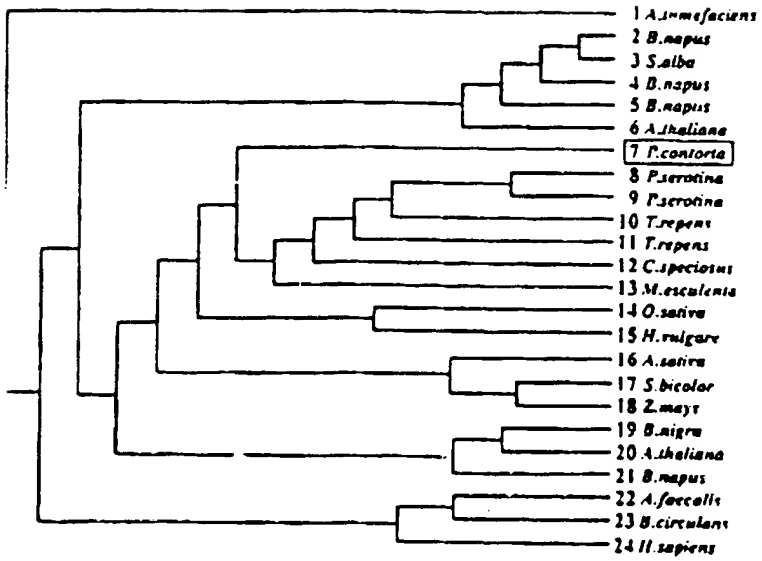
1 A.tumefaciens
2 B.napus
3 S.alba
4 B.napus
5 B.napus
6 A.thaliana
7 P.contorta
8 P.serotina
9 P.serotina
10 T.repens
11 T.repens
12 C.speciosus
13 M.esculenta
14 O.sativa
15 H.vulgare
16 A.sativa
17 S.bicolor
18 Z.mays
19 B.nigra
20 A.thaliana
21 B.napus
22 A.fatcalis
23 B.circulans
24 H.sapiens

Figure 2

```
CBGAA      MEVSVLMWVLLFYSLLGF-Q----------VTTARLDRN----N    29
L41869     MRSSPVL--LLVIALVAA-AHLAPLECDGPNPNPEIGNTGGLSRQ----G    43
J26025     T--KLGSLLLCALLLAGF-A-LTNSKAAKTDPPI--HCA-SLNRS----S    39
J39228     ----------LLLLGF-A-LANTNAARTDPPV--VCA-TLNRT----N    29
K56733     ----------LLSI-T-TTHIHAFKPLPISFDDFS-DLNRS----C    29
P26204     M--DFIVAIFALFVISSF-T-ITSTNA--VEASTLLDIG-NLSRS----S    39
Kg1986     MASKHSLHLFGLLIVFLV-S-LLLVLTNQATAFDGDFIPLNFSRS----Y    44
S35175     M----------LVLFIS-L-LALTRPAMGTDDDDDNIPDDFSRK----Y    33
J33817     MALLLASAINHTAHPAGLRSH---PNNESFSRHHLCSSPQNISKRRSNLS    47
A48860P    MAPLLAAAMNHAAAHPGLRSHLVGPNNESFSRHHLPSSSPQSSKRRCNLS    50
K78433     MA-LLCSALSNSTH-PSFRSH-IGANSENLW--HLSADPAQKSKRRCNLT    45
K89413     MVLQKLPLIGLLLLT------IVASPANAD-GPVCPPSNKLSRAS----    39
S52771P    M---KFPLLGLLLLVT------LVGSPTRAEEGPVCPKTETLSRAS----    37
Q00326     MKLLHGLALVFLLAAASCK----------ADEEITCEENNP    31
X79195     MKLL-GFALAILLW ATCK----------PEEEITCEENVP    30
L11454     MKLL-MLAFVFLLALATCK----------GDEFVCEENEP    29
LPH3HU     WEKF----------FSQPKFERDLF    15
A48969P    M----------1
Q08638     M----------1
S45675     VVP----------3

CBGAA      FPS----------DFMFGTASSAYQYEGAVRE    51
L41869     FPA----------GFVFGTAASAYQVEGMARQ    65
J26025     FDA----------LEPGFIFGTASAAYQFEGAAKE    64
J39228     FDT----------LFPGFTFGTATASYQLEGAANI    54
X56733     FA----------PGFVFGTASSAFQYEGAAFE    51
P26204     FP----------RGFIFGAGSSAYQFEGAVNE    61
X94986     FPD----------DFIFGTATSAYQIEGAANK    66
S35175     FPD----------DFIFGTATSAYQIEGEATA    55
J33817     FRPRAQTISSESAGIHRLSPWEIPRRDWFPPSFLFGAATSAYQIEGAWNE    97
A48860P    FTTRSARVGSQN-GVQMLSPSEIPQRDWFPSDFTFGAATSAYQIEGAWNE    99
K78433     LSSRAARISSALESAKQVKPWQVPKRDWFPPEFMFGAASAAYQIEGAWNE    95
K89413     ----------FPEGFLFGTATAAYQVEGAINE    61
S52771P    ----------FPEGFMFGTATASYQVEGAVNE    59
Q00326     FTCSNTDILSSKN----------FGKDFIFGVASSAYQIEGGR--    64
X79195     FTCSQTDRFNKQD----------FESDFIFGVASSAYQIEGGR--    63
L11454     FTCNQTKLFNSGN----------FEKGFIFGVASSAYQVEGGR--    62
LPH3HU     YHGT----------FRDDFLWGVSSSAYQIEGAWDA    41
A48969P    ----------SIHM----------FPSDFKWGVATAAYQIEGAYNE    27
Q08638     ----------NVKK----------FPEGFLWGVATASYQIEGSPLA    27
S45675     ---AAQQTATAPDAALT----------FPEGFLWGSATASYQIEGAAAE    39
                                       .* .* ..... * **
```

Figure 3

```
3GAA      DGKGPSTWDALTHM-PGRI-KDSSNGDVAVDQYHRYMEDIELMASLGLDA   99
41869     GGRGPCIWDAFVAI-QGMI-AGNGTADVTVDEYHRYKEDVGIMKNMGFDA  113
26025     DGRGPSIWDTYTHNHSERI-KDGSNGDVAVDQYHRYKEDVRIMKKMGFDA  113
39228     DGRGPSIWDAFTHNHPEKI-TDGSNGDVAIDQYHRYKEDVAIMKDMGLDA  103
56733     DGKGPSIWDTFTHKYPEKI-KDRTNGDVAIDEYHRYKEDIGIMKDMNLDA  100
26204     GGRGPSIWDTFTHKYPEKI-RDGSNADITVDQYHRYKEDVGIMKDQNMDS  110
94986     FGRGASVWDTFTHQYPERI-LDHSTGDVADGFYYRFKGDIQNVKNMGFNA  115
35175     KGRAPSVWDIFSKETPDRI-LDGSNGDVAVDFYNRYIQDIKNVKKMGFNA  104
33817     DGKGPSTWDHFCHNFPEWI-VDRSNGDVAADSYHMYAEDVRLLKEMGMDA  146
48860P    DGKGESNWDHFCHNHPERI-LDGSNSDIGANSYHMYKTDVRLLKEMGMDA  148
78433     GGKGPSSWDNFCHSHPDRI-MDKSNADVAANSYYMYKEDVRMLKEIGMDS  144
39413     TCRGPALWDIYCRRYPERC-NND-NGDVAVDFFHRYKEDIQLMKNLNTDA  109
52771P    GCRGPSLWDIYTKKFPHRV-KNH-NADVAVDFYHRFREDIKLMKKLNTDA  107
00326     -GRGVNVWDGFSHRYPEKAGSDLKNGDTTCESYTRWQKDVDVMGELNATG  113
79195     -GRGLNVWDGFTHRYPEKGGADLGNGDTTCDSYRTWQKDIDVMEELGVKG  112
11454     -GRGLNVWDSFTHRFPEKGGADLGNGDTTCDSYTLWQKDIDVMDELNSTG  111
PH3HU     DGKGPSIWDNFTHT-PGSNVKDNATGDIACDSYHQLDADLNMLRALKVKA   90
48969P    DGRGMSIWDTFAHT-PGKV-KNGDNGNVACDSYHRVEEDVQLLKDLGVKV   75
00638     DGAGMSIWHTFSHT-PGNV-KNGDTGDVACDHYNRWKEDIEIIEKLGVKA   75
45675     DGRTPSIWDTYART-PGRV-RNGDTGDVATDHYHRWREDVALMAELGLGA   87
```

```
3GAA      YRFSISWSRILPEGR--GEINMAGIEYYNNLIDALLQNGIQPFVTLFHFD  147
41869     YRFSISWSRIFPDGT--GKVNQEGVDYYNRLIDYMLQQGITPYANLYHYD  161
26025     YRFSISWSRVLPNGKVSGGVNEDGIKFYNNLINEILRNGLKPFVTIYHWD  163
39228     YRFSISWSRLLPNGTLSGGINKKGIEYYNNLTNELIRNGIEPLVTLFHWD  153
56733     YRFSISWPRVLPKGKLSGGVNREGINYYNNLINEVLANGMQPYVTLFHWD  150
26204     YRFSISWPRILPKGKLSGGINHEGIKYYNNLINELLANGIQPFVTLFHWD  160
94986     FRFLISWPRVIPSGTRREGINEQGIEFYNKVINEIINQGMEPFVTIFHWD  165
35175     FRMSISWSRVIPSGRRREGVNEEGIQFYNDVINEIISNGLEPFVTIFHWD  154
33817     YRFSISWPRILPKGTLAGGINEKGVEYYNKLIDLLLENGIEPYITIFHWD  196
48860P    YRFSISWPRILPKGTKEGGINPDGIKYYRNLINLLLENGIEPYVTIFHWD  198
78433     YRFSISWPRILPKGTLDGGINHEGIQYYNDLLDCLIENGIKPYITLFHWD  194
39413     FRMSIAWPRIFPHGRKEKGVSQAGVQFYHDLIDELIKNGITPFVTVFHWD  159
52771P    LRLSIAWPRIFPHGPMEKGNSKEGVQFYHDLIDELLKNDLTPLVTIFHWD  157
00326     YRFSFAWSRIIPKGKVSRGVNQGGLDYYHKLIDALLEKNITPFVTLFHWD  163
79195     YRFSFAWSRILPKGKRSRGINEDGINYYSGLIDGLIARNITPFVTLFHWD  162
11454     YRFSIAWSRLLPKGKRSRGVNPGAIKYYNGLIDGLVAKNMTPFVTLFHWD  161
PH3HU     YRFSISWSRIFPTGR-NSSINSHGVDYYNRLINGLVASNIFPMVTLFHWD  139
48969P    YRFSISWPRVLPQGT--GEVNRAGLDYYHRLVDELLANGIEPFCTLYHWD  123
00638     YRFSISWPRILPEGT--GRVNQKGLDFYNRIIDTLLEKGITPFVTIYHWD  123
45675     YRFSLAWPRIQPTGR--GPALQKGLDFYRRLADELLAKGIQPVATLYHWD  135
```

Figure 3

```
CBGAA      LPKALEDSYGGWLS---PQIINDFEAYAEICFRAFGDRVKYWATVNEPNL  194
L41869     LPLALHQQYLGWLS---PKIVGAFADYAEFCFKVFGDRVKNWFTFNEPRV  208
U26025     LPQALEDEYGGFLS---PNIVDHFRDYANLCFKKFGDRVKHWITLNEPYT  210
U39228     VPQALEEEYGGVLS---PRIVYDFKAYAELCYKEFGDRVKHWTTLNEPYT  200
X56733     VPQALEDEYRGFLG---PNIVDDFRDYAELCFKEFGDRVKHWITLNEPWG  197
P26204     LPQVLEDEYGGFLN---SGVINDFRDYTDLCFKEFGDRVRYWSTLNEPWV  207
X94986     TPQAIEDKYGGFLS---ANIVKDYREYADLLFERFGDRVKFWMTFNEPWS  212
S35175     TPQALQDKYGGFLS---RDIVYDYLQYADLLFERFGDRVKPWMTFNEPSA  201
U33817     TPQALVDAYGGFLDEED---YKDYTDFAKVCFEKFGKTVKNWLTFNEPET  243
A48860P    VPQALEEKYGGFLDKSHKSIVEDYTYFAKVCFDNFGDKVKNWLTFNEPQT  248
X78433     TPQALADEYKDFLD---RRIVKDYTDYATVCFEHFGDKVKNWFTFNEPHS  241
X89413     TPQDLEDEYGGFLSE---RIVKDFREYADFVFQEYGGKVKHWITFNEPWV  206
S52771P    MPADLEDEYGGFLSE---RVVPDFVEYANFTFHEYGDKVKNWITFNEPWV  204
Q00326     LPQTLQDEYEGFL---DRQIIQDFKDYADLCFKEFGGKVKHWITINQLYT  210
X79195     LPQSLQDEYEGFL---DRTIIDDFKDYADLCFERFGDRVKHWITINQLFT  209
L11454     LPQTLQDEYNGFL---NKTIVDDFKDYADLCFELFGDRVKNWITINQLYT  208
LPH3HU     LPQALQDI-GGW---ENPALIDLFDSYADFCFQTFGDRVKFWMTFNEPMY  185
A48969P    LPQALQDQ-GGW---GSRITIDAFAEYAELMFKELGGKIKQWITFNEPWC  169
Q08638     LPFALQLK-GGW---ANREIADWFAEYSRVLFENFGDRVKNWITLNEPWV  169
S45675     LPQELENP-GGW---PERPTAERFAEYAAIAADALGDRVKTWTTLNEPWC  181
            *    .     .                .  ..      *  .. * * *  .

CBGAA      FVPLGYTVGIFPPT-RCAAPHANPLCM-TGNCSSAEPYLAAHHVLLAHAS  242
L41869     VAALGYDNGFHAPG-RCSK------CP-AGGDSRTEPYIVTHNIILSHAA  250
U26025     FSSSGYAYGVHAPG-RCSA-WQKLNCT-GGN-SATEPYLVTHHQLLAHAA  256
U39228     ISNHGYTIGIHAPG-RCSS-WYDPTCL-GGD-SGTEPYLVTHNLLLAHAA  246
X56733     VSMNAYAYGTFAPG-RCSD-WLKLNCT-GGD-SGREPYLAAHYQLLAHAA  243
P26204     FSNSGYALGTNAPG-RCSA-SNVAK---PGD-SGTGPYIVTHNQILAHAE  251
X94986     LSGFAYDDGVFAPG-RCSS-WVNRQCR-AGD-SATEPYIVAHHLLLAHAA  258
S35175     YVGFAHDDGVFAPG-RCSS-WVNRQCL-AGD-SATEPYIVAHNLLLSHAA  247
U33817     FCSVSYGTGVLAPG-RCSP---GVSCAVPTGNSLSEPYIVAHNLLRAHAE  289
A48860P    FTSFSYGTGVFAPG-RCSP---GLDCAYPTGNSLVEPYTAGHNILLAHAE  294
X78433     FCGLGYGTGLHAPGARCSA---GMTCVIPEEDALRNPYIVGHNLLLAHAE  288
X89413     FLHAGYDVGKKAPG-RCSSYVNA---KCQDGRSGYEAYLVTHNLLISHAE  252
S52771P    FSRSAYDVGKKAPG-RCSPYIKDFGHLCQDGRSGFEAYVVSHNLLVSHAE  253
Q00326     VPTRGYAIGTDAPG-RCSP-MVDTKHRCYGGNSSTEPYIVAHNQLLAHAT  258
X79195     VPTRGYALGTDAPG-RCSQ-WVDK -RCYGGDSSTEPYIVAHNQLLAHAT  255
L11454     VPTRGYALGTDAPG-RCSP-KIDV--RCPGGNSSTEPYIVAHNQLLAHAA  254
LPH3HU     LAWLGYGSGEFPPGVK---------------DPGWAPYRIAHTVIKAHAR  220
A48969P    MAFLSNYLGVHAPGNK----------------DLQLAIDVSHHLLVAHGR  203
Q08638     VAIVGHLYGVHAPGMR----------------DIYVAFRAVHNLLRAHAR  203
S45675     SAFLGYGSGVHAPGRT----------------DPVAALRAAHHLNLGHGE  215
                .  *      .*.                  .      *     .*.
```

Figure 3

3-7

```
CBGAA      AVEKYREKYQKIQGGSIGLVISAPWYEPLENSP-EERSAVDRILSFNLRW  291
L41869     AVQRYREKYQPHQKGRIGILLDFVWYEPHSDTD-ADQAAAQRARDFHIGW  299
U26025     AVKLYKDEYQASQNGLIGITLVSPWFEPASEAE-EDINAAFRSLDFIFGW  305
U39228     AVKLYREKYQASQEGVIGITVVSHWFEPASESQ-KDINASVRALDFMYGW  295
X56733     AARLYKTKYQASQNGIIGITLVSHWFEPASKEK-ADVDAAKRGLDFMLGW  292
P26204     AVHVYKTKYQAYQKGKIGITLVSNWLMPLDDNSIPDIKAAERSLDFQFGL  301
X94986     AVKIYRENYQETQNGKIGITLFTYWFEPLSNSTD-DMQASRTALDFMFGL  307
S35175     AVHQYRKYYQGTQKGKIGITLFTFWYEPLSDSKV-DVQAAKTALDFMFGL  296
U33817     TVDIYNKYHKG-ADGRIGLALNVFGRVPYTNTFL-DQQAQERSMDKCLGW  337
A48860P    AVDLYNKHYKR-DDTRIGLAFDVMGRVPYGTSFL-DKQAEERSWDINLGW  342
X78433     TVDVYNKFYKG-DDGQIGMVLDVMAYEPYGNNFL-DQQAQERAIDFHIGW  336
X89413     AVEAYRK-CEKCKGGKIGIAHSPAWFEAHDLADSQDGASIDRALDFILGW  301
S52771P    AVDAFRK-CEKCKGDKIGIAHSPAWFEPEDVEGGQR--TVDRVLDFIMGW  300
Q00326     VVDLYRTKYKF-QKGKIGPVMITRWFLPFDESDPASIEAAERMNQFFHGW  307
X79195     VVDLYRTRYKY-QGGKIGPVMITRWFLPYDDTL-ESKQATWRAKEFFLGW  303
L11454     AVDVYRTKYKDDQKGMIGPVMITRWFLPFDHSQ-ESKDATERAKIFFHGW  303
LPH3HU     VYHTYDEKYRQEQKGVISLSLSTHWAEPKSPGVPRDVEAADRMLQFSLGW  270
A48969P    AVTLFRE---LGISGEIGIAPNTSWAVPYRRTKEDMEACLRVNGWSG-DW  249
Q08638     AVKVFRE---TVKDGKIGIVFNNGYFEPASEKEEDIRAVRFMHQFNNYPL  250
S45675     AVQALRDR--LPADAQCSVTLNIHHVRPLTDSEADADAVRRIDALAN-RV  262
           .      .        .              .

CBGAA      FLDPIVF-GDYPQEMR----------ERLGSRLPSISSELSAKLRGSFDY  330
L41869     FLDPITN-GRYPSSML----------KIVGNRLPGFSADESRMVKGSIDY  338
U26025     FMDPLTN-GNYPHLMR----------SIVGERLPNFTEEQSKLLKGSFDF  344
U39228     FMDPLTR-GDYPQSMR----------SLVKERLPNFTEEQSKSLIGSYDY  334
X56733     FMHPLTK-GRYPESMR----------YLVRKRLPKFSTEESKELTGSFDF  331
P26204     FMEQLTT-GDYSKSMR----------RIVKNRLPKFSKFESSLVNGSFDF  340
X94986     WMDPITY-GRYPRTVQ----------YLVGNRLLNFTEEVSHLLRGSYDF  346
S35175     WMDPMTY-GRYPRTMV----------DLAGDKLIGFTDEESQLLRGSYDF  335
U33817     FLEPW R-GDYPFSMR----------VSARDRVPYFKEKEQEKLVGSYDM  376
A48860P    FLEPW R-GDYPFSMR----------SLARERLPFFKDEQKEKLAGSYNM  381
X78433     FLEPMVR-GDYPFSMR----------SLVGDRLPFFTKSEQEKLVSSYDF  375
X89413     HLDTTTF-GDYPQIMK----------DIVGHRLPKFTTEQKAKLKASTDF  340
S52771P    HLDPTTY-GDYPQSMK----------DAVGARLPKFTKAQKAKLKGSADF  339
Q00326     YMEPLTK-GRYPDIMR----------QIVGSRLPNFTEEEAELVAGSYDF  346
X79195     FMEPLTK-GKYPYIMR----------KLVGNRLPKFNSTEARLLKGSYDF  342
L11454     FMGPLTE-GKYPDIMR----------EYVGDRLPEFSETEAALVKGSYDF  342
LPH3HU     FAHPIFRNGDYPDTMKWKVGNRSELQHLATSRLPSFTEEEKRFIRATADV  320
A48969P    YLDPIYF-GEYPKFM---LDWYENLGY-----KPPIVDGDMELIHQPIDF  290
Q08638     FLNPIYR-GDYPELV---LE-FAR-EY-----LPENYKDDMSEIQEKIDF  289
S45675     FTGPMLQ-GAYPEDL---VKDTAGLTD-----WSFVRDGDLRLAHQKLDF  303
                     * *.   .                                .
```

Figure 3

4-7

```
CBGAA       MGINHYTTLYATSTPPLSPDHTQ--YLYPDSRVYLTGERHGV-SIGERTG  377
L41869      VGINQYTSYYMKDPGAWNQTPVS--YQ-DDWHVGFVYERNGV-PIGPRAN  384
U26025      IGLNYYTTRYASNAPKITSVHA--SYITDPQVNAT-AELKGV-PIGPMAA  390
U39228      IGVNYYSARYASAYPEDYSIPTPPSYLTDAYVNVT-TELNGV-PIGPQAA  382
X56733      LGLNYYSSYYAAKAPRI--PNARPAIQTDSLINAT-FEHNGK-PLGPMAA  377
P26204      IGINYYSSSYISNAPSH--GNAKPSYSTNPMTNIS-FEKHGI-PLGPRAA  386
X94986      IGLQYYTSYYAKPNAPYDFNHIR--YLTDNRVTETPYDYNGN-LIGPQAY  393
S35175      VGLQYYTAYYAEPIPPVDPKFRR--YKTDSGVNATPYDLNGN-LIGPQAY  382
U33817      IGINYYTSTFSKHID-LSPNNSPVLNTDDAYASQETKGPDGN-AIGFPTG  424
A48860P     LGLNYYTSRFSKNID-ISPNYSPVLNTDDAYASQEVNGPDGK-PIGPPMG  429
X78433      VGINYYTSRFAKHID-ISPEFIPKINTDDVYSNPEVNDSNGI-PIGPDVG  423
X89413      VGLNYYTSVFSNHLEK--PDPSKPRWMQDSLITWESKNAQ-NYAIGSKPL  387
S52771P     VGINYYSSFYAKASEK--PDYRQPSWATDSLVEFEPKTVDGSVKIGSQPS  387
Q00326      LGLNYYVTQYAQPKPNPYPSETHTA-MMDAGVKLTYDNSRGEFL-GPLFV  394
X79195      LGLNYYVTQYAHALDPSPPEKL-TA-MTDSLANLTSLDANGQPP-GPPF-  388
L11454      LGLNYYVTQYAQNNQTIVPSDVHTA-LMDSRTTLTSKNATGHAP-GPPF-  389
LPH3HU      FCLNTY---YSRIVQHKTPRLNPPSYEDD--QEMAEEEDPSWPSTAMNRA  365
A48969P     IGINYYTSSMNRYNPGEAGGMLSSEAISMGAP---------------KTD  325
Q08638      VGLNYYSGHLVKFDPDAPAKV---SFVERDLP---------------KTA  321
S45675      LGVNYYSPTLVSEADGSGTHNSDGHGRSAHSPWPGADRVAFHQPPGETTA  353
              .. *

CBGAA       MDGLFVVPH----GIQKIVEYVKEFYDNPTII-IAENGYPESE--ESSST  420
L41869      SDWLYIVPW----GMNKAVTYVKERYGNPTMI-LSENGMDQP----GNVS  425
U26025      SGWLYVYP----KGIHDLVLYTKEKYNDPLIY-ITENGVDEFN--DPKLS  433
U39228      SDWLYVYP----KGLYDLVLYTKNKYNDPIMY-ITENGMDEFN--NPKIS  425
X56733      SSWLCIYP----QGIRKLLLYVKNHYNNPVIY-ITENGRNSST--INTV-  419
P26204      SIWIYVYPYMFIQEDFEIFCYILKINITILQFSITENGMNEFN--DATLP  434
X94986      SDWFYIFP----ESIRHLLNYTKDTYNDPVIY-ITENGVDNQN--NETEP  436
S35175      SSWFYIFP----KGIRHFLNYTKDTYNDPVIY-VTENGVDNYN--NESQP  425
U33817      NAWINMYP----KGLHDILMTMKNKYGNPPMY-ITENGMGDIDKGDLPKP  469
A48860P     NPWIYMYP----EGLKDLLMIMKNKYGNPPIY-ITENGIGDVDTKETPLP  474
X78433      MYFIYSYP----KGLKNILLRMKEKYGNPPIY-ITENGTADMDGWGNP-P  467
X89413      TAALNVYS----RGFRSLLKYIKDKYANPEIM-IMENGYGEELGASDSV-A  432
S52771P     TAKMAVYA----AGLRKLVKYIKDRYGNPEII-ITENGYGEDLGEKDTDHS  433
Q00326      EDKVNGNSYYYPKGIYYVMDYFKTKYGDPLIY-VTENG---FSTPSSENR  440
X79195      ----SKGSYYHPRGMLNVMEHFKTKYGDPLIY-VTENG---FSTSGGPIP  430
L11454      ----NAASYYYPKGIYYVMDYFKTTYGDPLIY-VTENG---FSTPGDE-D  430
LPH3HU      FPW----------GTRRLLNWIKEEYGDIPIY-ITENGVGLTNPNT----  400
A48969P     ICWE-----IYAEGLYDLLRYTADKYGNPTLY-ITENGA----CYNDGLS  365
Q08638      MGWE-----IVPEGIYWILKKVKEEYNPPEVY-ITENGA----AFDDVVS  361
S45675      MGWA-----VDPSGLYELLRRLSSDFPALPLV-ITENGA----AFHDYAD  393
                                              .***
```

Figure 3

```
CBGAA      LQENLNDVRRIRFHGDCLSYLSAAIKN-GSDVRGYFVWSLLDNFEWAFGY  469
L41869     IADGVHDTVRIRYYRDYITELKKAIDN-GARVAGYFAWSLLDNFEWRLGY  474
U26025     MEEALKDTNRIDFYYRHLCYLQAAIKK-GSKVKGYFAWSFLDNFEWDAGY  482
U39228     LEQALNDSNRIDYCYRHLCYLQEAIIE-GANVQGYFAWSLLDNFEWSEGY  474
X56733     -------TSRIPF-------------------------------------  425
P26204     VEEALLNTYRIDYYYRHLYYIRSAIRA-GSNVKGFYAWSFLDCNEWFAGF  483
X94986     IQDAVKDGFRIEYHRKHMWNALGSLKEYHVNLKGYFAWSYLDNFEWNIGY  486
S35175     IEEALQDDFRISYYKKHMWNALGSLKNYGVKLKGYFAWSYLDNFEWNIGY  475
U33817     V--ALEDHTRLDYIQRHLSVLKQSIDLGAD-VRGYFAWSLLDNFEWSSGY  516
A48860P    MEDALNDYKRLDYIQRHIATLKESIDLGSN-VQGYFAWSLLDNFEWFAGF  523
X78433     MTDPLDDPLRIEYLQQHMTAIKEAIDLGRRTLRGHFTWSLIDNFEWSLGY  517
X89413     AV-GTADHNRKYYLQRHLLSMQEAVCIDKVNVTGYFVWSLLDNFEWQDGY  490
S52771P    SV-ALNDHNRKYYHQRHLLSLHQAICEDKVNVTSYFVWSLMDNFEWLDGY  481
Q00326     -EQAIADYKRIDYLCSHLCFLRKVIKEKGVNVRGYFAWALGDNYEFCKGF  489
X79195     FTEAFHDYNRIDYLCSHLCFLRKAIKEKRVNVKGYFVWSLGDNYEFCNGY  480
L11454     FEKATADYKRIDYLCSHLCFLSKVIKEKNVNVKGYFAWSLGDNYEFCNGF  480
LPH3HU     -----EDTDRIFYHKTYINEALKAYRLDGIDLRGYVAWSLMDNFEWLNGY  445
A48969P    LDGRIHDQRRIDYLAMHLIQASRAIED-GINLKGYMEWSLMDNFEWAEGY  414
Q08638     EDGRVHDQNRIDYLKAHIGQAWKAIQE-GVPLKGYFVWSLLDNFEWAEGY  410
S45675     PEGNVNDPERIAYVRDHLAAVHRAIKD-GSDVRGYFLWSLLDNFEWAHGY  442
                 *  .                  .. ..  .. ..  .

CBGAA      TIRFGLYHVDFIS-DQKRYPKLSAQWFRQFLQHDDQGS------------  506
L41869     TARFGIVYVDF-N-TLKRYPKDSALWFKNMLSEKKRS-------------  509
U26025     TVRFGINYVDYND-NLKRHSKLSTYWFTSFLKKYERSTKEIQMFVESKLE  531
U39228     TVRFGINYVDYDN-GLKRHSKLSTHWFKNFLKRSSISKEKIRRCGNNNAR  523
X56733                                                         425
P26204     TVRFGLNFVD-----------------------------------------  493
X94986     TARFGLYYVDYNN-NLTRIPKDSAYWFKAFLN-PENITKTTRTVSWDSRK  534
S35175     TSRFGLYYVDYKN-NLTRYPKKSAHWFTKFLNISVNANNIYELTSKDSRK  524
U33817     TERFGIVYVDREN-GCERTMKRSARWLQEF------------NGAAKKVE  553
A48860P    TERYGIVYVDRNN-NCTRYMKESAKWLKQF------------NAAKKP--  558
X78433     LSRFGIVYIDRND-GCKRIMKKSAKWLKEF------------NGATKKLN  554
X89413     KNRFGLYYVDFKN-NLTRYEKESGKYYKDFLSQGVRPSALKKDE------  523
S52771P    TARFGLYYIDFQN-NLTRMEKESATCS---LNSSNRA-------------  514
Q00326     TVRFGLSYVNWEDL-DDPNLKESGKWYQRFIN------GTVKNAVKQDFL  532
X79195     TVRFGLSYVDFNNVTADRDLKASGLWYQSFLR---------DTTKNQDIL  521
L11454     TVRFGLSYVDFANITGDRDLKASGKWFQKFIN------VTDEDSTNQDLL  524
LPH3HU     TVKFGLYHVDFNNTNRPRTARASARYYTEVITNNG---------------  480
A48969P    GMRFGLVHVDYDTL--VRTPKDSFYWYKGVISRGWL-D------------  449
Q08638     SKRFGIVYVDYSTQ--KRIVKDSGYWYSNVVKNNGL-E----------.--  445
S45675     SKRFGAVYVDYPTG--TRIPKASARWYAEVARTGVLPT------------  478
                    .

CBGAA      -----------IRSSSS--I  513
L41869     --------------------  509
U26025     HQKFESQMMNKVQSSLAVVV  551
U39228     ARKFVYR------------I  531
X56733     --------------------  425
P26204     --------------------  493
X94986     AGKF-------------YIM  541
S35175     VGKF-------------YVM  531
U33817     NNKI---------LTPAGQLN  565
A48860P    SKKI---------LTPA----  566
X78433     NKILGASSCCSGVTHGGGTA  574
X89413     -------------------L  524
S52771P    --------------------  514
Q00326     RSSLSSQS-QKKRFADA---  548
X79195     RSSLPFKNGDRKSLT-----  536
L11454     RSSVSSKNRDRKSLADA---  541
LPH3HU     -------------------M  481
A48969P    -------------------L  450
```

Figure 3

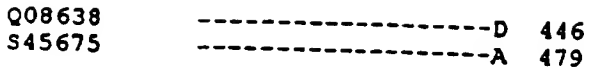

Figure 3

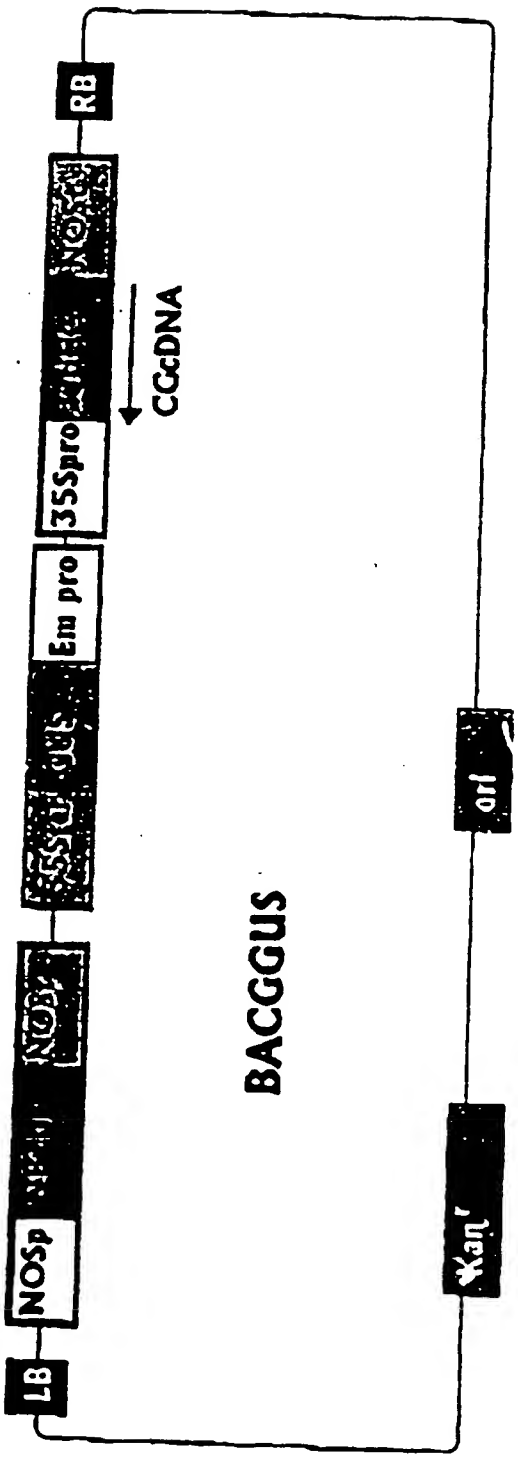

Figure 4